8-2016

Propensity score based methods for estimating the treatment effects based on observational studies.

Younathan Abdia

PROPENSITY SCORE BASED METHODS FOR ESTIMATING THE TREATMENT EFFECTS BASED ON OBSERVATIONAL STUDIES

By

Younathan Abdia
MS, The Ohio State University, 2011

A Dissertation
Submitted to the Faculty of the
School of Public Health and Information Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy in Biostatistics

Department of Bioinformatics and Biostatistics
University of Louisville
Louisville, KY 40202

August, 2016

PROPENSITY SCORE BASED METHODS FOR ESTIMATING THE TREATMENT EFFECTS BASED ON OBSERVATIONAL STUDIES

Submitted by

Younathan Abdia

A Dissertation Approved on

August 5th 2016

by the Following Dissertation Committee:

Maiying Kong, Ph.D
Dissertation chair

Somnath Datta, PhD
Dissertation Co-chair

K.B Kulasekera, Ph.D
Dissertation Co-chair

Maxwell Boakye, MD, MPH, MBA

Riten Mitra, PhD

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my advisors, Dr. Maiying Kong, Dr. Somnath Datta, and Dr. K. B. Kulasekera for their support and guidance throughout my study here at the University of Louisville. With their insightful guidance and unconditional support I am able to finish my research projects. I particularly like to thank Dr. Kong for her time to meet me every week, and for her patience and guidance through the last two years.

I would also like to thank Dr. Maxwell Boakye from the Department of Neurosurgery and Dr. Riten Mitra from the Department of Bioinformatics and Biostatistics for serving on my dissertation committee. I especially thank Dr. Boakye and Dr. Beatrice Ugiliweneza for providing the MarketScan data for my dissertation.

I would like to thank the Department of Bioinformatics and Biostatisitcs for the financial support during my study here at the University of Louisville. I would also like to thank to Mrs. Lynne Dosker for her administrative support which made my life easier and more enjoyable.

I would also like to thank my friends and peers: Chathura Siriwardhana, Jasmit Shah, Soutik Ghosal, Debamita Kundu, Joe Bible and Yubing Wang for their discussion on academic problems and for their friendship and all the fun we had together in the past.

I would like to thank my Church family and friends. Their prayer and love have always supported me. I would like to thank Lewellens, Brooks, Otts, Cecils, Grays, Mike Yeager, Frames, Paxhias, Rick Nutt and Mary Gene Boteler for being

wonderful friends and for constant care and love. I would also thank my community group members for it was wonderful to know all of them.

I am so grateful to God for He has given me wonderful parents and siblings. My parents have always supported and encouraged me. Their unconditional love has always been my strength. My family is always praying for me to get best in my life. Last but not the least, I am thankful to my wonderful and beautiful wife, Krona Daniel. Indeed, she is the best and the biggest blessing in my life. She has supported me so much throughout this program, without her this journey could be very difficult. She has done so many sacrifices for me and I will never forget them. I dedicate my dissertation to my parents, wife, and my siblings.

ABSTRACT

PROPENSITY SCORE BASED METHODS FOR ESTIMATING THE TREATMENT EFFECTS BASED ON OBSERVATIONAL STUDIES

Younathan Abdia

August 5th 2016

This dissertation consists of two interconnected research projects. The first project was a study of propensity scores based statistical methods for estimating the average treatment effect (ATE) and the average treatment effect among treated (ATT) when there are two treatment groups. The ATE is defined as the mean of the individual causal effects in the whole population, while ATT is defined as the treatment effect for the treated population. Propensity score based statistical methods, such as matching, regression, stratification, inverse probability weighting (IPW), and doubly robust (DR) methods were used to estimate the ATE and ATT. Simulation studies and case studies were conducted to examine the performances of propensity score based methods when propensity score was estimated using logistic regression and generalized boosted models (GBM). The aim of the second project is to develop generalized propensity score based statistical methods for estimating ATE when there are more than two treatment groups. The generalized propensity score was estimated using Multinomial logistic regression, random forests, and GBM. In addition, an adaptive optimal ensemble method was developed to estimate the generalized propensity score. Once the generalized propensity scores were obtained, IPW, stratification, and DR methods were used to estimate the ATE. Simulation studies were conducted to examine the performances of these different

generalized propensity score based methods. In addition, we applied these methods to examine the outpatients health care costs under different treatments for patients with spinal fusion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

In randomized control trials (RCT), the subjects are randomly assigned to different groups, say, treatment and comparator groups. In an RCT, it is generally assumed that there are no confounding baseline covariates, either measured or unmeasured (Austin, 2011). Thus, the treatment effect can be estimated directly by comparing outcomes between the treatment group and the comparator group (Austin, 2011). However, it is not always feasible to carry out an RCT due to ethical or practical reasons. As a result, a lot of data have been collected under natural settings such as registry data, claim data, and electronic clinical records where confounding may arise. Appropriate use of these data could provide valuable information to health care providers and policy makers. Unlike in an RCT, the treatment assignment is no longer random, and the treatment assignment may depend on the patient's characteristic covariates. Thus, the covariates may be related to treatment assignment as well as outcome variables. The difference of the outcome variable between treatment and comparator groups may be due to either the treatment or covariates. It would be improper to compare the outcomes between the treatment and comparator groups without accounting for the confounding of the covariates.

There are two commonly investigated quantities in the observed data under natural settings: the average treatment effect (ATE) and the average treatment effect among treated (ATT). Rosenbaum and Rubin (1983) introduced the concept of propensity score to help one make decisions based on observational data. The

propensity score is the probability of treatment assignment conditional on observed baseline covariates (Austin, 2011). According to Austin (2011), "the propensity score allows one to design and analyze an observational (nonrandomized) study, so that it mimics some of the particular characteristics of a RCT". Since the seminal work by Rosenbaum and Rubin (1983), many propensity score based methods have been proposed to make causal inferences based on observed data. These methods include matching (Rosenbaum and Rubin, 1983; Rosenbaum and Rubin, 1985), regression with propensity score as covariate (Rosenbaum and Rubin, 1983; Rosenbaum, 1987), stratification (Rosenbaum and Rubin, 1983; Rosenbaum and Rubin, 1984; Lunceford and Davidian, 2004), inverse probability of weighting (IPW) (Rosenbaum 1987; Lunceford and Davidian, 2004), and the doubly robust method (Lunceford and Davidian, 2004). However, the propensity score is generally unknown and is estimated using a logistic regression model (Rosenbaum and Rubin, 1983). Using logistic regression may lead to a biased estimate of propensity score if the model is misspecified. McCaffrey et al. (2004) suggested generalized boosted model (GBM) to estimate the propensity score. GBM is a nonparameteric technique; it selects the important covariates and their interactions, and may provide lower prediction errors. For GBM based propensity score estimates, only IPW methods for estimating ATT have been investigated (McCaffrey et al., 2004). To the best of our knowledge, the performance of regression and doubly robust methods based on GBM estimated propensity score have not been investigated in the literature. For Project 1, we provide an overall comparative study of the commonly used propensity score based methods for estimating ATT and ATE, when the propensity score are estimated by logistic regression and GBM, respectively.

Project 2 furthers our investigation to examine statistical methods for estimating generalized propensity score when there are multiple treatment groups. An adaptive ensemble method is developed to estimate the generalized propensity

score. Imbens (2000) extends the use of propensity score from two treatment groups to multiple treatment groups by introducing the generalized propensity score. The generalized propensity score is the conditional probability of receiving a particular level of a treatment given pre-treatment variables (Imbens, 2000). Imbens (2000) provides a theoretical framework for assessing treatment effects when multiple treatment groups are involved. Lechner (2001) outlines how to assess treatment effects (ATT and ATE) for multiple treatment groups using a matching method based on the generalized propensity score. The stratification (i.e., subclassification) method for multiple treatment groups has been developed by Zanutto et al. (2005). The regression adjustment and weighting methods for estimating ATE has been developed and examined by Feng et al. (2010).

The generalized propensity score is usually estimated by multinomial logistic regression. However, when the number of variables in the regression is large, variable selection and justification to add non-linear and interaction terms in the linear predictor for estimating the generalized propensity score could be challenging. McCaffrey et al. (2004) propose a non-parametric method (i.e. GBM) to estimate the generalized propensity score. GBM selects important covariates and their interaction terms. The GBM method has been applied to estimate the propensity score for two groups (McCaffrey et al., 2004) and has been extended to estimate the generalized propensity score for multiple treatment groups (McCaffrey et al., 2013).

In recent years, machine learning techniques have been used to estimate propensity score. Setoguchi et al. (2008) compare the propensity score estimates using logistic regression, classification and regression trees (CART), pruned CART and neural networks, and concluded that the propensity score estimated by neural networks provides the least biased estimates for ATE and ATT. Lee et al. (2010) used logistic regression, CART, pruned CART, bagging, random forests, and GBM

to estimate the propensity score, and concluded that the bagging, random forests and GBM models perform excellently in covariate balance and treatment effect estimation. Lee et al. (2010) recommend GBM and random forests to estimate propensity scores. To the best of our knowledge, only GBM has been used to estimate the generalized propensity score. Other machine learning methods, such as CART, prune CART, bagging and random forests are only applied to estimate the treatment effect between two groups. Although it is straight forward to extend these methods to estimate the generalized propensity score when multiple treatment groups are involved, it is probably difficult to provide a universal answer on which method is optimal to estimate the generalized propensity score. The general role of the generalized propensity score is to balance the covariates. McCaffrey et al. (2013) propose using the absolute standardized mean difference (ASMD) and Kolmogorov-Smirnov statistic to assess whether each covariate adjusted by the generalized propensity score is balanced. Based on the work by Datta et al. (2010), an adaptive optimal ensemble method, which uses bagging and rank aggregation, is constructed to estimate generalized propensity scores. The adaptive optimal ensemble method balances covariates. In Project 2, extensive simulation studies were carried out to examine the performances of the ensemble based techniques in estimating the treatment effects when multiple treatment groups are involved. A case study is carried out to examine the cost among different treatments for patients with spinal fusion.

CHAPTER 2

PROPENSITY SCORES BASED METHODS FOR ESTIMATING AVERAGE TREATMENT EFFECT AND AVERAGE TREATMENT EFFECT AMONG TREATED: A COMPARATIVE STUDY

2.1 Introduction

In randomized control trials (RCT), the subjects are randomly assigned to different groups, say, treatment and comparator groups. In an RCT, it is generally assumed that there are no confounding baseline covariates, either measured or unmeasured (Austin, 2011). Thus, the treatment effect on an outcome can be estimated directly by comparing outcomes between the treatment group and the comparator group (Austin, 2011). However, it is not always feasible to carry out an RCT due to ethical or practical reasons. As a result, a lot of data have been collected under natural settings where confounding may arise, such as registry data, claim data, and electronic clinical records. Appropriate use of these data could provide valuable information to health care providers and policy makers. Unlike in an RCT, the treatment assignment is no longer random, and the treatment assignment may depend on the patient's characteristic covariates. For example, a doctor may make a treatment choice based on the patient's age and current health conditions. Thus, covariates may be related to treatment assignment as well as the outcome variables. The difference of the outcome variable between treatment and comparator groups may be due to either treatment or covariates. It would be improper to compare the outcome between the treatment and comparator groups without accounting for the

confounding of the covariates.

There are two commonly investigated quantities in the observed data under natural settings: the average treatment effect (ATE) and the average treatment effect among treated (ATT). Rosenbaum and Rubin (1983) introduced the concept of propensity score to help one make decisions based on observational data. The propensity score is the probability of treatment assignment conditional on observed baseline covariates (Austin, 2011). According to Austin (2011), "the propensity score allows one to design and analyze an observational (nonrandomized) study, so that it mimics some of the particular characteristics of a RCT". Since the seminal work by Rosenbaum and Rubin (1983), many propensity score based methods have been proposed to make causal inferences for observational studies. These methods include matching (Rosenbaum and Rubin, 1983 and 1985; Rosenbaum, 1989), regression with propensity score as covariate (Rosenbaum and Rubin, 1983; Rosenbaum, 1987), stratification (Rosenbaum and Rubin,1983 and 1984; Lunceford and Davidian, 2004), inverse probability weighting (IPW) (Rosenbaum, 1987; Lunceford and Davidian, 2004; Austin, 2012), and the doubly robust method (Lunceford and Davidian, 2004). However, the propensity score is generally unknown and is estimated using a logistic regression model (Rosenbaum and Rubin, 1983). Using logistic regression may lead to a biased estimator of propensity score if the model is misspecified. McCaffrey et al. (2004) suggested generalized boosted regression (GBM) to estimate the propensity score. GBM is a nonparameteric technique; it selects the important covariates and their interactions, and may provide lower prediction errors. For GBM based propensity score, only IPW methods have been investigated in the literature (McCaffrey et al., 2004; Austin, 2012). To the best of our knowledge, the performance of regression and doubly robust methods based on GBM estimated propensity score have not been investigated in the literature. In this article, we provide an overall comparative study of the commonly used propen-

sity score based methods for estimating ATT and ATE, where the propensity scores are estimated by logistic regression and GBM. As commonly used in the literature, the average treatment effect (ATE) is defined as the mean of the individual causal effects in the whole population, while average treatment effect among treated (ATT) is defined as the mean of the individual causal effect in the treated population.

The rest of the chapter is organized as follows. Section 2.2 is an overall review on the basic assumptions for causal inference and the estimation methods for the propensity score. In Section 2.3, the propensity score based methods for estimating ATT are presented, while in Section 2.4, the propensity score based methods for estimating ATE are presented. Propensity score is also called a balancing score, which balances the covariates between treatment and comparator groups. In Section 2.5, the criterion for assessing covariate balance for each method is presented. In Section 2.6, extensive simulations are carried out to examine the performance of these methods. In Section 2.7, two case studies are carried out to illustrate how to apply these methods. The last section is devoted to a discussion.

2.2 Basic assumptions for causal inference and methods for the estimating propensity score

We present a few formal definitions to make the concepts clear. Let Z be a binary indicator variable: $Z = 1$ if a subject is in the treatment group and $Z = 0$ if a subject is in the comparator group. Every subject in the population has two potential outcomes (McCaffrey et al., 2004): the potential outcome when the subject were in the treatment group (say, Y_1), and the potential outcome when the subject were in the comparator group (say, Y_0). The observed outcome is

$$Y = ZY_1 + (1 - Z)Y_0 = \begin{cases} Y_0, & \text{if } Z{=}0, \\ Y_1, & \text{if } Z{=}1. \end{cases} \tag{2.1}$$

A subject in a study can only take one of the potential outcomes depending on his/her group assignment. The other potential outcome is often called the counterfactual outcome. Let X denote the covariate which may impact the selection of the treatment and the outcome variable. We assume that all relevant covariates are observed. The basic assumptions for causal inference are:

(i) *Temporality*: the treatment selection Z must occur before outcome;

(ii) *Overlap*:

$$0 < P[Z = 1|X] < 1. \tag{2.2}$$

That is, each subject in the study has the potential to be treated with the treatment or comparator (Williamson et al., 2011);

(iii) *Strongly ignorable treatment assumption (SITA)*: the potential outcomes (Y_0, Y_1) are independent of the treatment selection given the observed covariates X:

$$(Y_1, Y_0) \perp\!\!\!\perp Z|X; \tag{2.3}$$

(iv) *Stable unit treatment value assumption (SUTVA)*: the potential outcomes of one subject is not affected by the potential outcome of another subject.

Equation (2.3) under the assumption (iii) is also known as the conditional independence assumption (CIA) (Angrist and Pischke, 2009). Under assumptions (i) - (iv), one may replace the counterfactual outcome by the observed outcomes in other subjects which have the same covariates but in the alternative group. However, due to multi-dimensionality, specially in high-dimensional covariate cases, seeking subjects in the alternative group with same covariate values becomes a difficult task.

According to Rosenbaum and Rubin (1983), the propensity score is defined as

$$e(X) = P[Z = 1|X]. \tag{2.4}$$

Rosenbaum and Rubin (1983) proved that the assumptions (ii) and (iii) imply that

$$0 < P[Z = 1|e(X)] < 1, \tag{2.5}$$

and

$$(Y_1, Y_0) \perp\!\!\!\perp Z|e(X). \tag{2.6}$$

Thus, one may replace the counterfactual outcome by the observed outcome in other subjects with the same propensity score but in the alternative group. The propensity score is the conditional probability of receiving a treatment assignment with given covariates X (Rosenbaum and Rubin, 1983; Rosenbaum, 2009), and formally propensity score is expressed in equation (2.4). In an RCT, the propensity score is usually known. For example, in a two arm RCT with equal sample size, $e(X) = P[Z = 1|X] = \frac{1}{2}$ for all values of covariates X. The propensity score is also called a balancing score. That is, if treated and control subjects have the same propensity score, then the distribution of X for treated subjects and that for the comparators are the same. Mathematically speaking, it implies that $Z \perp\!\!\!\perp X|e(X)$. The propensity score has played an important role in causal inference. The estimation of the propensity score is usually carried out using the logistic regression technique (Rosenbaum and Rubin, 1983). In recent years, nonparameteric methods, such as generalized boosted method (GBM), have been applied to estimate the propensity score to alleviate potential issues caused by model misspecification.

2.2.1 Logistic regression to estimate propensity score

In a non-randomized study, the propensity score function is unknown. Logistic regression has been widely applied to estimate the propensity score. A logistic

regression has the following form:

$$log\left\{\frac{e(X)}{1-e(X)}\right\} = X\beta. \tag{2.7}$$

The predicted value $\hat{e}(X)$ is the estimated propensity score. In the above model, X usually includes all available covariates. Nonlinear terms such as quadratic and interaction terms, are not usually included in the model. It also becomes difficult to include nonlinear terms when the number of covariates is large.

2.2.2 Generalized boosting method to estimate propensity score

Recently, generalized boosting method (GBM) has been proposed for estimating the propensity score and is found to be able to improve the prediction accuracy (McCaffrey et al., 2014; Burgette et al., 2015). GBM is an automated data-adaptive algorithm, which uses regression trees as weak predictors and captures nonlinear and interactive effects of the covariates (Hastie et al., 2009). GBM uses the "forward stagewise additive algorithm" to estimate the propensity score by modeling $g(X) = logit\big(e(X)\big) = log\big(e(X)/(1-e(X))\big)$, where X is the covariates and $e(X)$ is the propensity score defined in equation (4). GBM begins with a single regression tree, for example, taking $\hat{g}(x) = log\big(\bar{z}/(1-\bar{z})\big)$ as an initial estimate of $logit\big(e(X)\big)$, where $\bar{z}$ is the mean of the treatment indicator variable in the sample. GBM adds a simple regression tree (say, $\hat{h}(x)$) to the currently fitted $logit\big(e(X)\big)$ (say $\hat{g}(x)$) to obtain a better fit. The added simple regression tree is obtained by fitting the residuals of the currently estimated $logit\big(e(X)\big)$ versus X (Burgette et al., 2015). The $logit\big(e(X)\big)$ is updated by $\hat{g}(x) + \lambda\hat{h}(x)$, where λ is known as the shrinkage factor or the learning rate. Generally the smaller the λ, the smoother the estimated propensity score. The shrinkage value is usually less than 1. McCaffery et al. (2004) recommended using 0.001 or 0.005 for the shrinkage parameter. It has been seen that the computational time increases almost linearly with the reciprocal

of the shrinkage factor (McCaffery et al., 2004). Usually, the propensity score estimates are obtained until a prespecified maximum number of iterations is reached (McCaffrey et al., 2004; Burgette et al., 2015; Hastie et al., 2009). We used the *gbm* function from the *gbm* package in R to construct the GBM models of interaction depth one, two, and three with a shrinkage factor at 0.05. The GBM model with the maximum likelihood function was selected to estimate the propensity score.

2.3 Propensity score based methods for estimating ATT

Average treatment effect among treated (ATT) is the treatment effect for the treated population. Mathematically,

$$\begin{aligned} ATT = E[Y_1 - Y_0|Z=1] &= E_X\big(E(Y_1 - Y_0|X, Z=1)\big) \\ &= E_X\big(E(Y_1|X, Z=1) - E(Y_0|X, Z=1)\big). \end{aligned} \tag{2.8}$$

Under SITA,

$$E(Y_0|X, Z=1) = E(Y_0|X, Z=0),$$

this implies,

$$ATT = E_X\big(E(Y_1|X, Z=1) - E(Y_0|X, Z=0)\big). \tag{2.9}$$

Under the basic assumptions for causal inference, from the definition for propensity score (Rosenbaum and Rubin, 1983), the following equation holds (Rosenbaum and Rubin, 1983; Angrist and Pischke, 2009):

$$ATT = E_{e(X)}(E(Y_1|e(X), Z=1) - E(Y_0|e(X), Z=0)). \tag{2.10}$$

Many statistical methods have been proposed to estimate ATT. The fundamental idea is that the subjects with covariates X (or $e(X)$) in the treatment group are compared with those in the comparator group with the same covariates X (or $e(X)$). To estimate ATT, the distribution X is based on the treated population. In the

following section, the commonly used propensity score based statistical methods for estimating ATT are presented.

2.3.1 Optimal pair matching within propensity score calipers

Matching is the most popular technique for estimating ATT. It is generally assumed that the comparator group has more subjects than the treatment group. Each subject in the treatment group is matched with a subject in the comparator group based on their covariates and the propensity score. The propensity score is used to set up a caliper where the difference of the propensity scores of the two matched subjects is within the caliper. Two commonly used matching techniques are greedy matching and optimal matching (Rosenbaum, 1989), each based on the distance between the covariates, such as the Mahalanobis distance, of two subjects (Rosenbaum, 2009). However, the distance includes an additional penalty term if the propensity score of the two subjects is outside the caliper (Rosenbaum, 1989). That is, the distance between the i^{th} subject in the treated group and j^{th} subject in the comparator group is defined as

$$Mahalanobis\ dist(X_i, X_j) + \delta I[|\hat{e}(X_i) - \hat{e}(X_j)| > caliper],$$

where δ is a very large positive number, and I is an indicator variable. In optimal matching each treated subject is matched with a comparator subject to minimize the total distance. In greedy matching, each treated subject is matched with a comparator, where the distance is the smallest without considering the overall matches (Stuart, 2010). Gu and Rosenbaum (1993) concluded that optimal matching and greedy matching perform equally in terms of creating groups with good balance, but the optimal matching does reduce the distance within pairs. In this project, we apply the optimal matching to obtain a matching comparator subject for each treated subject. Once we have all pairs matched, two sample paired t-test is applied

to draw inference. Since each treated subject is matched with a comparator subject, this approach is appropriate for estimating ATT.

2.3.2 Propensity score adjusted regression method

Rosenbaum and Rubin (1983) provided a theortical framework for using the propensity score adjusted regression to estimate ATT. A regression model of the following form is used:

$$E[Y|Z,X] = \beta_0 + \beta_1 e(X) + \beta_2 Z + \beta_3 e(X)Z. \tag{2.11}$$

After fitting the above model, ATT is estimated by the treatment effect at the sample mean of the propensity score at the treated group (Williamson et al., 2011). That is, $\hat{\mu}_{ATT,Reg} = \hat{\beta}_2 + \hat{\beta}_3 \overline{e(X_T)}$, where $\overline{e(X_T)}$ is the sample mean of the propensity score of subjects with $Z = 1$. The variance can be obtained by $Var(\hat{\mu}_{ATT,Reg}) = (1, \overline{e(X_T)}) Var(\hat{\beta}_2, \hat{\beta}_3)(1, \overline{e(X_T)})^t$. The estimator $\hat{\mu}_{ATT,Reg}$ used here is in line with the approach by Imbens (2004), where the two potential outcomes for each subject (say, i^{th} subject with covariate X_i) are estimated, the difference of the two estimated potential outcomes is calculated $\big(i.e., \hat{\beta}_2 + \hat{\beta}_3 e(X_i)\big)$, and the mean of differences over all treated subjects is taken as the estimate for ATT. This resulting quantity is exactly the same as $\hat{\mu}_{ATT,Reg}$.

2.3.3 Propensity score based stratification method

Stratification could be used to estimate ATT (Williamson et al., 2011). In stratification, subjects are first ranked according to their estimated propensity score, then the strata are created according to cut off points defined by the quantiles of the estimated propensiy scores (Austin, 2011). Subjects with similar propensity score are placed into one stratum. It has been shown that stratification with five

strata based on the quantiles removes 90% of the bias in estimating treatment effect (Austin, 2011; Rosenbaum and Rubin, 1984). ATT can be estimated by the sum of the weighted treatment effect in each stratum, where the weight of the stratum is the proportion of the treated subjects in the stratum over all treated subjects in the sample. That is,

$$\hat{\mu}_{ATT,Strat} = \sum_{k=1}^{K} \frac{N_{T_k}}{N_T} \hat{\tau}_k,$$

where K is the number of strata, N_{T_k} is the number of the treated subjects in the k^{th} stratum, N_T is the total number of treated subjects in the sample, and $\hat{\tau}_k$ is the estimated treatment effect for the k^{th} stratum. The quantity τ_k is usually estimated by the difference of the mean of the treated subjects versus the comparator subjects in the k^{th} stratum. The variance of $\hat{\tau}_k$ is estimated by the pooled sample variances of the two samples within k^{th} stratum, and an estimated variance for $\hat{\mu}_{ATT,Strat}$ can be obtained by $\sum_{k=1}^{K} (\frac{N_{T_k}}{N_T})^2 \widehat{Var}(\hat{\tau}_k)$.

2.3.4 Inverse probability weighted method

The inverse probability weighted (IPW) method is to weight the treated and comparator observations to make them representative of the population of interest. To estimate ATT, the weight for a treated subject is taken as one, and the weight for a comparator subject is defined as $\frac{e(X)}{1-e(X)}$ (Imbens, 2004). Suppose, there are n subjects in the sample. Denote X_i, Z_i, and Y_i, respectively, as the observed covariates, treatment assignment, and outcome for the i^{th} subject ($i = 1, \cdots, n$). The IPW estimator for ATT Imbens (2004) is defined as:

$$\hat{\mu}_{ATT,IPW} = \frac{\sum_{i=1}^{n} Z_i Y_i}{\sum_{i=1}^{n} Z_i} - \frac{\sum_{i=1}^{n} (1 - Z_i) Y_i e(X_i) / \big(1 - e(X_i)\big)}{\sum_{i=1}^{n} (1 - Z_i) e(X_i) / \big(1 - e(X_i)\big)}. \tag{2.12}$$

Since the propensity score, $e(X_i)$, is unknown it is replaced by its estimate $\hat{e}(X_i)$. To estimate the standard error, Imbens (2004) recommended the bootstrap method,

stating that it leads to a valid standard error and a confidence intervals for the IPW estimate for ATT.

2.4 Propensity score based methods for estimating ATE

The average treatment effect (ATE) is the treatment effect in the entire population, which is defined as:

$$ATE = E(Y_1 - Y_0) = E_X(E(Y_1 - Y_0)|X) = E_X(E(Y_1|X) - E(Y_0|X)). \quad (2.13)$$

Under the SITA, ATE can also be written as:

$$ATE = E_X(E(Y_1|X, Z = 1) - E(Y_0|X, Z = 0)). \quad (2.14)$$

Due to Rosenbaum and Rubin (1983), under SITA, equations (2.13) and (2.14) hold if X is replaced by $e(X)$. The statistical methods for estimating ATE are parrallel to those for estimating ATT. However, the optimal matching for each treated subject is not applicable for estimating ATE.

2.4.1 Propensity score adjusted regression method for ATE

The propensity score adjusted regression model is the same as what is presented in Section 2.3.2. However, since ATE is the average treatment effect for the entire population, the mean of propensity score should be calculated over the entire sample (Williamson et al., 2011). That is, $\hat{\mu}_{ATE,Reg} = \hat{\beta}_2 + \hat{\beta}_3\overline{e(X)}$, where $\overline{e(X)}$ is the mean of propesnity score over the entire sample. The variance can be obtained by $Var(\hat{\mu}_{ATE,Reg}) = (1, \overline{e(X)})Var(\hat{\beta}_2, \hat{\beta}_3)(1, \overline{e(X)})^t$. This approach is in line with the approach recommended by Imbens (2004), where one first calculates the difference of two potential outcomes for i^{th} subject (i.e., $\hat{\beta}_2 + \hat{\beta}_3\hat{e}(X_i)$), and then takes the mean over the entire sample (say, $\hat{\beta}_2 + \hat{\beta}_3\overline{e(X)}$) as the estimate for ATE.

2.4.2 Propensity score based stratification method for ATE

The propensity score based stratification method for ATE is parallel to that for ATT. The only difference is how the weight is assigned for each stratum. The weight for each stratum should be assigned as the proportion of the number of subjects in k^{th} stratum over the total number of subjects in the entire population. That is,

$$\hat{\mu}_{ATE,Strat} = \sum_{k=1}^{K} \frac{N_k}{N} \hat{\tau}_k.$$

Here N_k is the number of subjects in the k^{th} stratum, and N is the number of subjects in the entire sample. The variance of $\hat{\tau}_k$ is estimated by the pooled sample variances of the two samples within k^{th} stratum, and an estimated variance for $\hat{\mu}_{ATE,Strat}$ can be obtained by $\sum_{k=1}^{K} (\frac{N_k}{N})^2 \widehat{Var}(\hat{\tau}_k)$.

2.4.3 Inverse probability weighted method for ATE

The inverse probability weighted (IPW) method for ATE is to weight each observation to make it representative of the entire population. The weight for a treated subject is $\frac{1}{e(X)}$ and for a comparator subject is $\frac{1}{1-e(X)}$. The following IPW estimator for ATE was proposed by Rosenbaum (1998):

$$\hat{\mu}_{ATE,IPW} = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i Y_i}{e(X_i)} - \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - Z_i) Y_i}{1 - e(X_i)}. \tag{2.15}$$

Again, $e(X)$ is generally unknown, and $e(X_i)$ in the equation (2.15) is replaced by its estimate. When $e(X)$ is estimated by logistic regression, an explicit variance formula for $\hat{\mu}_{ATE,IPW}$ has been given by Lunceford and Davidian (2004). However, when $e(X)$ is estimated by GBM, there is no explicit formula for variance estimation, and the bootstrap method is recommended. Since the weights in equation (2.15) do not add to one for each group for a given sample, Imben (2004) proposed a

normalized estimator for ATE where the weights add to one for each group. We call it the normalized IPW method (IPWN), which can be written as:

$$\hat{\mu}_{ATE,IPWN} = \frac{\sum_{i=1}^{n} Z_i Y_i / \hat{e}(X_i)}{\sum_{i=1}^{n} Z_i / \hat{e}(X_i)} - \frac{\sum_{i=1}^{n} (1 - Z_i) Y_i / [1 - \hat{e}(X_i)]}{\sum_{i=1}^{n} (1 - Z_i) / [1 - \hat{e}(X_i)]}. \tag{2.16}$$

The variance of the estimator can be obtained using the bootstrap method whether the propensity score is estimated by logistic regression or GBM.

2.4.4 Doubly robust estimator for ATE

The doubly robust (DR) estimator proposed by Robins et al. (1994) is an amendment to the IPW method: it combines the propensity score regression and outcome regression. The DR estimator remains consistent if either the propensity score model or the outcome regression model is specified correctly (Robins et al., 1994; Lunceford and Davidian, 2004). The DR estimate for ATE is given by

$$\begin{aligned} \hat{\mu}_{ATE,DR} = & \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i Y_i - (Z_i - \hat{e}(X_i)) m_1(X_i, \hat{\alpha}_1)}{\hat{e}(X_i)} \\ & - \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - Z_i) Y_i + (Z_i - \hat{e}(X_i)) m_0(X_i, \hat{\alpha}_0)}{1 - \hat{e}(X_i)}. \end{aligned} \tag{2.17}$$

Here $\hat{e}(X_i)$ is the estimator of the propensity score for the i^{th} subject, and $m_z(X, \alpha_z)$ is the outcome regression model for regressing Y on X for group $Z = z$. A variance estimate can be obtained using the bootstrap method.

2.5 Assessment of covariate balance

Propensity score is known as the balancing score (Rosenbaum and Rubin, 1983), which means that the distribution of each covariate is the same between the treatment and comparator groups, with a given propensity score (Harder et

al., 2010). To quantify the balance of each covariate between the treatment and comparator groups, the absolute standardized mean difference (ASMD) (McCaffrey et al., 2013) has been used. Let us assume that there are J covariates, denoted by $X_{\cdot j}$ $(j = 1, \cdots, J)$. Let $\bar{X}_{\cdot j}^{(T)}$ and $\bar{X}_{\cdot j}^{(C)}$ be the mean of the j^{th} covariate in treatment group and comparator group, respectively, and let us denote $SD_j^{(T)}$ and SD_j as the standard deviation of the j^{th} covariate in the treatment group and in the entire sample, respectively. The ASMD for j^{th} covariate when estimating ATT is defined as

$$ASMD_j^{(ATT)} = \frac{|\bar{X}_{\cdot j}^{(T)} - \bar{X}_{\cdot j}^{(C)}|}{SD_j^{(T)}}, \tag{2.18}$$

and the ASMD for j^{th} covariate when estimating ATE is defined as

$$ASMD_j^{(ATE)} = \frac{|\bar{X}_{\cdot j}^{(T)} - \bar{X}_{\cdot j}^{(C)}|}{SD_j}. \tag{2.19}$$

Generally, an ASMD value of greater than 0.20 is considered as problematic and an evidence of imbalance of a covariate; it could be a potential source of bias (McCaffrey et al., 2013). Equations (2.18) and (2.19) could be used for the original observed data. When there are unbalanced covariates, propensity score adjusted comparisons are formed, which depend on the different methods for estimating ATT and ATE. In the following two subsections, we present how to assess covariate balance when ATT or ATE is estimated.

2.5.1 Assessment covariate balance when estimating ATT

The different methods for estimating ATT (i.e., matching, stratification and IPW) form different comparison groups. When matching is used, each subject in the treatment group is matched with a subject in the comparator group, which has the smallest Mahalanobis distance within the caliper. To examine whether the j^{th} covariate is balanced, the mean of the j^{th} covariate in the comparator group in

equation (2.18) is based on all matched subjects. All the other terms in equation (2.18) stay the same.

To assess the covariate balance for stratification, the absolute mean difference in each stratum is calculated, then the average of the absolute mean differences among different strata is set as the numerator in equation (2.18) . That is,

$$ASMD_j^{(ATT,Strata)} = \frac{\frac{1}{K}\sum_{k=1}^{K} |\bar{X}_{\bullet j,k}^{(T)} - \bar{X}_{\bullet j,k}^{(C)}|}{SD_j^{(T)}}, \tag{2.20}$$

where $\bar{X}_{\bullet j,k}^{(T)}$ ($\bar{X}_{\bullet j,k}^{(C)}$)is the mean of the j^{th} variable in the treatment group (in the comparator group) in the k^{th} strata, and $SD_j^{(T)}$ is the standard deviation of the j^{th} variable in the treatment group as defined in equation (2.18).

To assess the covariate balance for IPW method, the mean for comparator group in equation (2.18) is taken as the weighted group mean, the weights are the same as those in estimating ATT using IPW method. That is,

$$ASMD_j^{(ATT,IPW)} = \frac{|\bar{X}_{\bullet j}^{(T)} - \bar{X}_{\bullet j}^{(C,IPW)}|}{SD_j^{(T)}}, \tag{2.21}$$

where $\bar{X}_{\bullet j}^{(C,IPW)} = \frac{\sum_{i=1}^{n}(1-Z_i)X_{ij}e(X_i)/(1-e(X_i))}{\sum_{i=1}^{n}(1-Z_i)e(X_i)/(1-e(X_i))}$, where X_{ij} is the i^{th} observed value for j^{th} covariate (McCaffrey et al., 2013; Ridgeway et al., 2015). $\bar{X}_{\bullet j}^{(T)}$ and $SD_j^{(T)}$ are defined as in equation (2.18). One may calculate the ASMD upon using different estimating methods. The methods which provide balance of covariates may result in appropriate estimate for ATT.

2.5.2 Assessment of covariate balance when estimating ATE

When estimating ATE, the initial covariate balance could be evaluated by equation (2.19). A value of ASMD greater than 0.20 may indicate an unbalanced covariate. The ASMD statistic after stratification is given by

$$ASMD_j^{(ATE,Strata)} = \frac{\frac{1}{K}\sum_{k=1}^{K} |\bar{X}_{\bullet j,k}^{(T)} - \bar{X}_{\bullet j,k}^{(C)}|}{SD_j}, \tag{2.22}$$

where $\bar{X}^{(T)}_{\bullet,j,k}$ and $\bar{X}^{(C)}_{\bullet,j,k}$ and SD_j are defined as before.

The assessment for covariate balance for IPW related methods is similar to equation (2.19). However, the group mean is taken as the weighted group mean. The weight for a subject in the treatment group equals the reciprocal of its propensity score, while the weight for a subject in the comparator group equals the reciprocal of 1 minus its propensity score. The estimating method with good balance of covariates is more likely to provide an appropriate estimate for ATE.

2.6 Simulation studies

In this section, we conducted simualtions to examine the performance of different methods for estimating ATT and ATE. In our simulations we considered 10 continuous covariates, $X = (X_{\bullet 1}, \cdots, X_{\bullet 10})'$. These covariates can impact the treatment selection and the outcome variable. However, from the assumptions previously stated in Section 2.2, we made the treatment selection and the outcome variable independent given X. The models we considered for our simulation study are summarized in Table 2.1.

Table 2.1: Simualtion models

True propensity score model (T.PS)	$logit(p_{treat}) = \alpha_{0,treat} + \alpha_1 X_{\bullet 1} + \alpha_2 X_{\bullet 2} + \alpha_3 X_{\bullet 3} + \alpha_{11} X^2_{\bullet 1} + \alpha_{22} X^2_{\bullet 2} + \alpha_{23} X_{\bullet 2} X_{\bullet 3} + \alpha_{123} X_{\bullet 1} X_{\bullet 2} X_{\bullet 3}.$
True outcome regression model (T.OR)	$Y = \beta_0 + \beta_1 X^2_{\bullet 1} Z + \beta_2 X_{\bullet 4} Z + \beta_3 X_{\bullet 1} X_{\bullet 4}(1-Z) + \beta_4 X_{\bullet 5}(1-Z) + \epsilon.$
False propensity score model (F.PS)	$logit(p_{treat}) = \alpha_0 + \sum_{j=1}^{10} \alpha_j X_{\bullet j}.$
False outcome regression model (F.OR)	$Y = \beta_0 + \mu Z + \sum_{j=1}^{10} \beta_j X_{\bullet j} + \epsilon.$

In the sequel, we refer to the true propensity score model as T.PS, the true outcome regression model as T.OR, the false propensity score model as F.PS, and the false outcome regression model as F.OR.

The T.PS model above specifies the relationship of the treatment selection

probability to covariates. T.PS was used to generate the treatment probability, p_{treat}, for each subject with covariate value X, and the treatment assignment Z was generated from a Bernoulli distribution with parameter p_{treat}. In T.OR, ϵ was taken to be a normal random variable with zero mean and variance σ^2. The variance σ^2 is chosen by setting the signal to noise ratio (SNR) as 50, where the SNR is defined as $SNR = \frac{Var(E(Y|X))}{\sigma^2}$. In practice it is quite common that all covariates are included in the propensity score model and the outcome regression model in a linear fashion in F.PS and F.OR, as shown in Table 2.1.

2.6.1 Simulation scenarios

We considered two simulation scenarios to estimate ATT and ATE using different estimation methods defined in sections 2.3 and 2.4. Propensity score was estimated using logistic regression and GBM, defined in Section 2.2. The two scenarios were:

- *Scenario 1:* $X_{\cdot 1}, X_{\cdot 2}, \cdots, X_{\cdot 10}$ follow a multivariate normal distribution with zero mean, unit variance and correlation coefficient 0.5 for all distinct pairs of variables.
- *Scenario 2:* $X_{\cdot 1}, X_{\cdot 2}, \cdots, X_{\cdot 10}$ are independent normal random variables with zero mean and variance 1.

For the sake of brevity, we describe the simulation steps under Scenario 1 only:

Step 1. Generate 1000 realizations of $X_{\cdot 1}, X_{\cdot 2}, \cdots, X_{\cdot 10}$ to simulate 1000 observations for the covariates $X = (X_{\cdot 1}, \cdots, X_{\cdot 10})'$. Here $X_{\cdot 1}, X_{\cdot 2}, \cdots, X_{\cdot 10}$ were generated from a multivariate normal distribution with zero mean, unit variance and correlation coefficient 0.50 for all distinct pairs of variables.

Step 2. Using the 1000 realizations of $(X_{\cdot 1}, \cdots, X_{\cdot 10})'$ generated in Step 1, calculate 1000 treatment selection probability, p_{treat}, using the T.PS setting $(\alpha_1, \alpha_2, \alpha_3, \alpha_{11}, \alpha_{22}, \alpha_{23}, \alpha_{123}) = \big(log(1.25), log(1.5), log(1.75), log(1.25), log(1.5), log(1.75), log(2)\big)$. The coefficients $log(1.25)$, $log(1.5)$, $log(1.75)$, and $log(2)$ are considered, respectively, as the weak, moderate, strong, and very strong effect in the treatment selection model. The concept of assigning $\alpha' s$ as weak, moderate, strong, and very strong was given by Austin (2014). Here $\alpha_{0,treat}$ was selected such that approximately one third of the subjects were assigned to the treatment group.

Step 3. Generate 1000 realizations for the treatment assignment, Z, from Bernoulli distribution with parameters p_{treat} above.

Step 4. Generate 1000 realizations for the response variable using the T.OR model in Table 2.1, based on the 1000 realizations of $(X_{\cdot 1}, X_{\cdot 2}, \cdots, X_{\cdot 10})'$ in Step 1 and 1000 realizations of Z in Step 2. In the T.OR model $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = (0, 2, 3, 2, -4)$, and the σ^2 was set so that SNR = 50.

Step 5. Estimate the propensity score using T.PS, F.PS, and GBM, respectively. For GBM, the number of trees and the interaction depths were selected using a cross-validation method.

Step 6. Estimate the ATT, ATE, and their standard errors using the methods described in Sections 2.3 and 2.4.

Repeat Step 1 to Step 6 1000 times.

We report the mean of 1000 estimates of ATT and ATE by each method (see the column "Estimates" in Table 2.2), the mean of 1000 estimates of ATT and ATE by each method minus the true value (see the column "Bias" in Table 2.2), the average of the 1000 estimated standard errors (see the column "SE" in Table

2.2), the empirical standard deviation of the estimates (see the column "ESE" in Table 2.2) and the root mean square error (RMSE) of the estimates (see the column "RMSE" in Table 2.2). The empirical standard deviation is defined as the standard deviation of the 1000 estimated treatment effects under each method. The average of the 1000 standard errors being close to the empirical standard error indicates that the variance estimation for the method is appropriate. The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\mu}_i - \mu)^2},$$

where $\hat{\mu}_i$ denotes the estimated treatment effect (estimate of ATT or ATE) and μ is the true value of ATT or ATE. The true value for ATE can be calculated from the underlying setting. In the current setting the true ATE is 1. However, the underlying true ATT is not known. The true ATT can be obtained from the simulated data. That is, for each simulated data, based on the underlying outcome regression model, the two potential outcomes (Y_0, Y_1) were generated, the difference between the two potential outcomes as the treatment effect for the particular subject was calculated. The average of the treatment effects of the subjects in the treatment group, was considered as the sample specific value of ATT. The average of the 1000 sample specific values of ATT is considered as the true value of ATT, which is 3.962 for simulation Scenario 1.

To examine how well the propensity score performs in balancing each covariate between treatment and comparator groups, we calculated the balancing scores (i.e. ASMD) for each simulated data. The balancing scores include the balancing score of the original simulated data, and the balancing scores of each covariate upon using different propensity score based methods. The boxplots of the 1000 balancing scores for estimating ATT are presented in Figure 2.1, and those for ATE are in Figure 2.2.

We carried out the same simulations with a sample size of 5000 to examine the performances of each method when the sample size becomes larger. The results are reported in Table 2 under the column "Sample Size = 5000". For Scenario 2, the simulation steps remain the same except in Step 1, realizatons of 10 covariates were generated from an independent normal random variables with zero mean and variance 1. The simulation results are reported in Table 3, and the boxplots of the balancing scores are presented in Figure A1 for ATT and Figure A2 for ATE in the appendix.

We also carried out the simulations when the sample size was small, say 100, for Scenarios 1 and 2. To prevent strata from having a frequency of zero due to small sample size (n=100), the number of strata was reduced to four when ATT and ATE were estimated using the stratification. The simulation results for Scenarios 1 and 2 are reported respectively in Tables 2.2 and 2.3 in the supplementary material. All the simulations were carried out using the R statistical software version 3.1.2. For computation efficieny simulations were carried out in a cluster computing environment using parallel computing.

2.6.2 Simulation results

In this subsection, we discuss simulation results for the scenarios outlined in the previous subsection. We first examine the balancing of covaraites. Figure 2.1 represents the boxplots of the covariate balancing scores for estimating ATT when the covaraites are normally distributed with non zero correlations. Panel A indicates that all the covariates are unbalanced because the ASMD scores are greater than 0.20. Panels B1, C1, and D1 indicate that upon matching, the covariates are still unbalanced (i.e., greater than 0.20) irrespective to the propensity score estimation technique. Panels B2, C2, and D2 indicate that balancing of covariates

upon stratification is achieved by using the true propensity score model and GBM method but not by using false propensity score model. Panels B3, C3, and D3 indicate that balancing of covariates upon IPW is achieved only by using T.PS model, not by F.PS and GBM. The assessments of covariate balancing for Scenario 2 for estimating ATT are presented in Figure A1 in the appendix. $X_{\bullet 2}$ and $X_{\bullet 3}$ are not balanced for the simulated data. However, upon using different propensity score based methods, the covariate balancing is achieved for all but F.PS based stratification method.

The assessment of the covariates balancing for estimating ATE is presented in Figure 2.2, where the covariates are normally distributed with non zero correlations. Panel A indicates that all of the covariates are unbalanced (i.e., ASMDs are greater than 0.20). However, when the true propensity score model is applied, the covariate balances are achieved (Panels B1 and B2). When the false propensity score model is applied, the covariate $X_{\bullet 2}$ is not balanced for both stratification and IPW (Panels C1 and C2). When GBM is applied to estimate the propensity score, all the balancing scores (i.e., ASMD scores) upon using stratification are close to 0.20 (Panel D1), and the balancing scores (i.e., ASMD scores) for $X_{\bullet 3}$ upon using IPW are above 0.20 (Panel D2), indicating unbalance of covariate $X_{\bullet 3}$.

In Scenario 1 the true value for ATT is 3.962. By examining the simulation results (Table 2.2), we concluded that the ATT estimations based on stratification and IPW were close to the true value of ATT, whereas, matching and regression seemingly resulted in biased estimates for ATT. As the sample size increases from 1000 to 5000, we observed that the bias of the average of the estimates for ATT decreases when it is estimated using stratification and IPW methods. The estimates for ATT based on matching and regression were far from the true ATT. When the sample size is reduced to 100 the results exhibited similar behavior but with increased standard errors. Thus, matching and regression may not be suitable to

estimate ATT when covariates are correlated.

The true value of ATT under Scenario 2 is 2.485. In Scenario 2, as observed in Table 2.3, it appears that matching, regression, stratification and IPW provide the average of ATT estimates close to 2.485, regardless of the propensity score estimation method. When the sample size increases from 1000 to 5000, the bias of the average of the estimates for ATT decreases regardless of the propensity score estimation method, and the standard errors also decreased as expected. When the sample size was reduced to 100 all methods provided estimates close to the true value, with biases within 0.1. This suggests that all methods were appropriate for ATT estimation in Scenario 2. An additional simulation with settings similar to Leacy and Stuart (2014), was carried out and the results were reported in the appendix Table A1, which shows that the stratification and IPW with misspecified propensity score may result in ATT estimates with larger biases.

The true ATE under Scenario 1 is 1. The simulation results for ATE under Scenario 1 are presented in Table 2.2 under the row "ATE". When propensity score was estimated using the T.PS model, all methods except regression provided ATE estimates near the true value. When propensity score was estimated using F.PS, all methods, except DR, provided estimates with biases larger than 0.35. The estimates for ATE under DR method using T.OR and F.PS was close to the true ATE. When propensity score was estimated with GBM, the estimates for ATE based on stratification, IPWN, and DR with T.OR were close to the underlying true value 1. However, when the propensity score was estimated using GBM, the DR with F.OR provided an average of estimates at 1.569, which was 0.569 larger than the underlying ATE. The bias decreased to 0.345 as the sample size was increased to 5000, and the bias increased to 0.948 as the sample size was decreased to 100. The regression method for ATE gives a biased estimates, regardless of the sample size and propensity score estimation method.

The true value of ATE under Scenario 2 is 2. The simulation results for ATE under Scenario 2 are presented in Table 2.3 under the row "ATE". When the propensity score was estimated using the true logistic regression model (T.PS), the ATE estimates using the regression, stratification, IPW, IPWN, and DR were close to 2. When the propensity score was estimated using the falsely specified logistic regression model (F.PS), all the methods except DR with T.OR provided estimates for ATE ranging from 2.488 to 2.711. The results remained similar when the sample size was either increased to 5000 or decreased to 100. When the propensity score was estimated using the GBM method, all estimates for ATE were close to 2, with a bias within 0.24. The bias was reduced when the sample size was increased to 5000.

In reality, when estimating propensity score using logistic regression, the functional form in the linear predictor is unknown. Therefore, it is difficult to assess the accuracy of the parametric propensity score model. Propensity score estimated by GBM is an alternative method to alleviate model misspecification issues. From simulation results under both scenarios, stratification, IPWN and DR seem to provide more accurate estimates for ATE and thus are preferred. Stratification and IPW methods are recommended for estimating ATT. In addition, for all methods, the average of the standard errors was close to the empirical standard error, indicating that the variance estimation technique for each method was appropriate.

2.7 Case studies

2.7.1 Case study for ATT

The data set from the Lalonde's National Supported Work Demonstration was used to demonstrate the estimation results for ATT using different methods.

The data set was obtained from the R package *twang* (Imbens, 2000; Ridgeway et al, 2015). In the Lalonde dataset, the variable '*treat*' was a binary variable: 1 for treatment and 0 for comparator. The treatment 1 indicates that the subject was a part of the National Supported Work, and 0 indicates that the subject was from the Current Population Survey (Ridgeway et al., 2015). There were a total of 614 subjects in the dataset. Of these 614 subjects, 185 were in the treatment group and the rest in the comparator group. The covariates for estimating the propensity score were age, race, education (number of years), marriage status, earnings in 1974, and earnings in 1975, as illustrated in the package *twang* (Ridgeway et al., 2015). The objective of the National Supported Work Demonstration was to determine whether there was an increase in earnings for the year 1978. We used different propensity score based methods to estimate ATT, where the propensity score was estimated using both logistic regression and GBM. The assessment of covariate balance is reported in the appendix Figure A3, and the results are reported in Table 2.4. From Table 2.4, it is seen that the ATT estimates using the matching method were similar whether the propensity score was estimated using logistic regression or GBM. The ATT estimates using matching were close to IPW, where the propensity score was estimated by GBM. The ATT estimates using IPW and GBM were similar to those reported in Ridgeway et al. (2015). The estimates based on regression and stratification were drastically different compared to those based on the matching method.

From the simulation results in Section 2.6, it was inferred that ATT was less biased when the propensity score was estimated using GBM, for both matching and IPW. Also, from the results in Section 2.6, regression and stratification methods were found to be rather different from the true ATT. IPW provided biased results when the propensity score was not correctly specified. Therefore, in this case study for knowing the above observations regarding IPW and GBM,we conclude that the

estimated increased difference in earnings for year 1978 was 628, with a standard error 941, suggesting that the increase between the two groups was not statistically significant.

2.7.2 Case study for ATE

To examine different methods for estimating ATE, we used the Lindner data set provided in the *twang* package (Ridgeway et al., 2015). The Lindner data set consisted of 996 patients, treated at the Lindner Center in the Christ Hospital (Cincinnati, OH) in 1997. The patients were given percutaneous coronary intervention (PCI). One of the outcome variables was the cost for the first six months after treatment. The treatment variable was *abcix*, where 0 indicated that patient was in PCI group and 1 indicated that patient was in PCI treatment with additional treatment abciximab. Lindner data set includes the following covariates: (i) acutemi; 1 indicates recent acute myocardinal infraction, 0 otherwise; (ii) Left ventricle ejection fraction (a percentage between 0 and 90) (iii) Number of vessels involved in initial PCI (iv) Stent indicator variable for whether coronary stent was inserted or not (v) Diabetic indicator variable for whether the subject was diagnosed with diabetes or not (vi) Height and (vii) Gender.

The assessment of covariate balance is reported in the appendix in Figure A4, and the ATE estimates based on different methods are reported in Table 2.5. Based on the simulation results in Section 2.6, when the propensity score model was misspecified, all methods may provide biased results. Simulation studies from Section 2.6 also suggests stratification, IPWN, and DR may provide less biased estimates when the propensity score was estimated using GBM. The results from the case study align with our simulation results. When the propensity score was estimated using GBM, the ATE estimates from stratification, IPWN, and DR were close to

each other. When the propensity score was estimated using logistic regression, the ATE estimates from the three methods were much larger than the ATE estimates with the propensity score estimated using GBM. Drawing conclusions based on the ATE estimates from stratification, IPWN, and DR with propensity score estimated by GBM, we concluded that the cost of the first six months after treatment was roughly between 799 and 862 dollars with standard error between 759 and 918. Hence, the cost difference does not appear to be significantly different from zero.

2.8 Discussion

In this comparative study, we considered different statisitcal methods for estimating treatment effects. Based on our simulations and case studies, the estimates for ATT or ATE may vary greatly from one method to another. When the propensity score model is specified correctly, the regression method for both ATT and ATE may result in biased estimates, whereas stratification method provides reasonable estimates for ATE and ATT. When the propensity score model is misspecified, all methods except DR are biased for estimating ATT and ATE.

GBM provides an alternative approach for estimating the propensity score. When the propensity score was estimated using GBM, the resulting IPW estimates for ATT are comparable with those obtained from a correct specification of the propensity score model, and the resulting stratification, IPWN, and DR estimates for ATE are all comparable with those obtained under the correct specification of the propensity score model. We concluded that the IPW method using the GBM estimated propensity score may provide appropriate estimates for ATT and the stratification, IPWN, and DR using the GBM approach for propensity score are likely to provide appropriate estimates for ATE.

In this article, we investigated the causal inference when two groups are in-

volved. It is also important to make causal inference when multiple groups are involved. Some theoretical results on this area have been established recently (Imbens, 2000; Lechner, 2001; Imai and Van Dyk, 2004), and implemented to estimate ATT using GBM (McCaffrey et al., 2013) or using multinomial logistic regression (Feng et al., 2012). A thorough investigation on estimating both ATT and ATE for multiple group comparisons will have a great value. In this article, we have applied logistic regression and GBM to estimate the propensity score. Other methods, such as classification and regression trees (CART), pruned CART, bagged CART and random forests could also have been used to estimate propensity scores (Westreich et al., 2010). A data-driven ensemble classifier may improve the overall performance for causal inference, which is currently under investigation by our team.

Austin (2012) carried out an extensive simulation study to examine the performance of tree-based G-computation method for directly estimating ATE, where the tree-based ensemble outcome regression models were constructed for each treatment group, the predicted outcomes under treatment and comparator were respectively calculated for each subject, and the average of the differences of the predicted outcomes between treatment and comparator was the estimate of ATE. Austin (2012) compared the tree based G-computational method with the IPW method, where the propensity score was estimated using the tree-based methods including boosted regression tree. The results in Austin (2012) indicate that G-computation method has a superior performance in the majority of the simulated scenarios. G-computation method may deserve further investigation when multiple groups are involved.

Table 2.2: Simulation results for estimating ATT and ATE, where the covariates are dependently normally distributed, PS is estimated using the true logistic regression (T.PS), false logistic regression (F.PS), and GBM. The underlying ATE is 1 and ATT is 3.962.

		Sample size = 1000					Sample size = 5000					Sample size = 100				
Estimator	Method	Estimates	Bias	Std. Error	ESE	RMSE	Estimate	Bias	Std. Error	ESE	RMSE	Estimate	Bias	Std. Error	ESE	RMSE
ATT	Matching(T.PS)	3.239	0.723	0.362	0.360	0.830	3.299	0.663	0.160	0.157	0.692	3.074	0.888	1.138	1.121	1.429
	Matching(F.PS)	3.498	0.464	0.360	0.359	0.606	3.545	0.417	0.159	0.155	0.456	3.272	0.690	1.130	1.112	1.308
	Matching (GBM)	3.297	0.665	0.255	0.392	0.772	3.359	0.603	0.113	0.181	0.640	2.779	1.183	1.252	1.236	1.710
	Regression, $\overline{e(X_T)}$ (T.PS)	5.414	1.452	0.440	0.559	1.533	5.375	1.413	0.195	0.246	1.424	5.161	1.199	1.449	1.649	2.038
	Regression, $\overline{e(X_T)}$ (F.PS)	5.749	1.787	0.333	0.535	1.842	5.766	1.804	0.146	0.228	1.808	5.248	1.287	1.209	1.541	2.007
	Regression, $\overline{e(X_T)}$ (GBM)	6.402	2.440	0.524	0.864	2.574	5.926	1.964	0.208	0.351	1.956	5.561	1.600	1.783	2.737	3.168
	Stratification (T.PS)	3.732	0.230	0.442	0.468	0.533	3.693	0.269	0.194	0.120	0.344	3.669	0.102	1.640	1.626	1.651
	Stratification (F.PS)	3.982	0.020	0.366	0.439	0.439	3.946	0.016	0.161	0.188	0.190	3.852	0.293	1.414	1.396	1.399
	Stratification (GBM)	4.348	0.386	0.560	0.511	0.631	4.163	0.201	0.208	0.205	0.273	3.929	0.032	1.987	2.080	2.079
	IPW (T.PS)	3.891	0.071	0.557	0.731	0.737	3.909	0.053	0.270	0.327	0.327	3.764	0.198	1.386	1.499	1.511
	IPW (F.PS)	4.072	0.110	0.356	0.368	0.377	4.055	0.093	0.345	0.521	0.525	3.978	0.016	1.305	1.196	1.196
	IPW (GBM)	3.423	0.539	0.328	0.357	0.659	3.668	0.294	0.162	0.176	0.359	3.289	0.672	1.064	1.034	1.657
ATE	Regression $\overline{e(X)}$ (T.PS)	1.373	0.373	0.341	0.342	0.486	1.386	0.386	0.152	0.124	0.412	1.191	0.191	1.131	0.931	0.949
	Regression $\overline{e(X)}$ (F.PS)	1.770	0.770	0.296	0.307	0.840	1.808	0.808	0.130	0.137	0.825	1.614	0.614	1.067	1.031	1.200
	Regression $\overline{e(X)}$ (GBM)	1.432	0.432	0.372	0.338	0.541	1.422	0.422	0.156	0.143	0.444	1.037	0.037	1.285	1.277	1.277
	Stratification (T.PS)	1.038	0.038	0.363	0.305	0.307	1.038	0.038	0.160	0.126	0.134	1.032	0.032	1.295	1.037	1.037
	Stratification (F.PS)	1.352	0.352	0.304	0.295	0.468	1.335	0.335	0.133	0.128	0.366	1.466	0.446	1.188	1.066	1.163
	Stratification (GBM)	1.032	0.032	0.466	0.346	0.347	1.037	0.037	0.171	0.124	0.129	1.153	0.153	1.650	1.501	1.508
	IPW (T.PS)	0.988	0.012	0.481	1.065	1.065	1.040	0.040	0.272	1.969	1.969	0.932	0.068	2.448	1.320	1.321
	IPW (F.PS)	2.027	1.027	0.666	0.724	1.266	2.073	1.073	0.276	0.317	1.125	2.037	1.037	1.457	3.292	3.449
	IPW (GBM)	0.846	0.154	0.177	0.224	0.277	0.888	0.112	0.119	0.101	0.153	0.908	0.092	0.572	0.761	0.766
	IPW N (T.PS)	0.942	0.058	0.557	0.527	0.529	0.972	0.028	0.345	0.374	0.374	0.935	0.065	1.386	1.134	1.135
	IPW N (F.PS)	1.460	0.460	0.471	0.530	0.709	1.512	0.512	0.222	0.239	0.572	1.417	0.417	1.442	1.481	1.537
	IPW N (GBM)	1.060	0.060	0.276	0.254	0.315	0.985	0.015	0.120	0.108	0.109	1.338	0.338	0.906	0.888	0.950
	DR (T.PS, F.OR)	1.041	0.041	0.574	1.158	1.159	1.043	0.043	0.310	1.089	1.089	1.198	0.198	1.296	1.461	1.474
	DR (F.PS, T.OR)	0.987	0.013	0.220	0.223	0.223	0.994	0.006	0.098	0.095	0.095	0.980	0.020	0.703	0.732	0.732
	DR (F.PS, F.OR)	1.073	0.073	0.637	0.633	0.639	1.136	0.136	0.290	0.266	0.302	1.143	0.143	1.817	2.151	2.154
	DR (GBM, T.OR)	1.009	0.009	0.207	0.213	0.213	1.002	0.002	0.093	0.095	0.095	0.941	0.059	0.642	0.624	0.627
	DR (GBM, F.OR)	1.569	0.569	0.303	0.273	0.623	1.345	0.345	0.123	0.110	0.360	1.948	0.948	1.005	0.892	1.301

Note: T.OR indicates true outcome regression; F.OR indicates false outcome regression.

Table 2.3: Simulation results for estimating ATT and ATE, where the covariates are indenpendently normally distributed, PS is estimated using the true logistic regression (T.PS), false logistic regression (F.PS), and GBM. The underlying ATE is 2 and ATT is 2.485.

		Sample size = 1000					Sample size = 5000					Sample size = 100				
Estimator	Method	Estimates	Bias	Std. Error	ESE	RMSE	Estimate	Bias	Std. Error	ESE	RMSE	Estimate	Bias	Std. Error	ESE	RMSE
ATT	Matching(T.PS)	2.505	0.020	0.337	0.334	0.335	2.485	0.000	0.150	0.151	0.151	2.613	0.128	1.079	1.131	1.137
	Matching(F.PS)	2.492	0.007	0.337	0.365	0.365	2.487	0.002	0.150	0.157	0.156	2.562	0.077	1.083	1.140	1.142
	Matching (GBM)	2.499	0.014	0.343	0.375	0.375	2.494	0.009	0.165	0.165	0.165	2.560	0.075	1.101	1.162	1.164
	Regression, $\overline{e(X_T)}$ (T.PS)	2.510	0.025	0.390	0.424	0.425	2.488	0.003	0.172	0.185	0.185	2.551	0.066	1.317	1.466	1.467
	Regression, $\overline{e(X_T)}$ (F.PS)	2.479	0.006	0.328	0.426	0.426	2.484	0.001	0.146	0.180	0.180	2.554	0.069	1.120	1.1254	1.255
	Regression, $\overline{e(X_T)}$ (GBM)	2.542	0.057	0.450	0.577	0.578	2.490	0.005	0.180	0.222	0.222	2.489	0.005	1.509	1.786	1.785
	Stratification (T.PS)	2.505	0.020	0.374	0.412	0.413	2.486	0.001	0.165	0.176	0.176	2.578	0.093	1.343	1.498	1.500
	Stratification (F.PS)	2.486	0.001	0.333	0.370	0.370	2.485	0.000	0.148	0.159	0.159	2.584	0.099	1.154	1.159	1.163
	Stratification (GBM)	2.522	0.037	0.425	0.446	0.445	2.489	0.004	0.168	0.173	0.173	2.487	0.002	1.674	1.876	1.874
	IPW (T.PS)	2.474	0.011	0.515	0.663	0.663	2.482	0.003	0.270	0.327	0.327	2.559	0.074	1.445	1.527	1.528
	IPW (F.PS)	2.488	0.003	0.343	0.355	0.354	2.485	0.000	0.152	0.150	0.150	2.530	0.045	1.303	1.173	1.174
	IPW (GBM)	2.505	0.020	0.312	0.322	0.322	2.493	0.008	0.143	0.155	0.155	2.512	0.027	0.787	0.812	0.812
ATE	Regression $\overline{e(X)}$ (T.PS)	2.054	0.054	0.341	0.293	0.299	2.056	0.056	0.152	0.130	0.144	1.980	0.019	1.126	1.033	1.033
	Regression $\overline{e(X)}$ (F.PS)	2.488	0.488	0.316	0.283	0.567	2.487	0.487	0.141	0.124	0.508	2.476	0.476	1.060	0.955	1.066
	Regression $\overline{e(X)}$ (GBM)	2.025	0.025	0.359	0.300	0.301	2.020	0.020	0.152	0.124	0.126	2.321	0.321	1.257	1.204	1.246
	Stratification (T.PS)	2.011	0.011	0.354	0.302	0.302	2.000	0.000	0.156	0.134	0.134	2.012	0.012	1.216	1.078	1.078
	Stratification (F.PS)	2.507	0.507	0.320	0.283	0.584	2.501	0.501	0.142	0.125	0.521	2.550	0.550	1.107	1.052	1.187
	Stratification (GBM)	1.970	0.030	0.406	0.321	0.322	1.948	0.016	0.158	0.116	0.126	2.311	0.311	1.564	1.507	1.538
	IPW (T.PS)	2.000	0.000	0.373	0.704	0.703	1.995	0.005	0.183	0.260	0.260	1.965	0.035	2.448	1.202	1.202
	IPW (F.PS)	2.711	0.711	0.327	0.345	0.794	2.690	0.690	0.143	0.146	0.710	2.712	0.712	6.936	1.552	1.708
	IPW (GBM)	1.892	0.100	0.154	0.242	0.264	1.919	0.081	0.069	0.104	0.131	2.074	0.074	0.536	0.824	0.827
	IPW N (T.PS)	1.995	0.005	0.385	0.380	0.489	1.995	0.005	0.270	0.200	0.220	1.993	0.007	1.445	1.137	1.136
	IPW N (F.PS)	2.558	0.558	0.303	0.319	0.646	2.574	0.574	0.134	0.137	0.569	2.540	0.540	1.336	1.165	1.283
	IPW N (GBM)	2.184	0.184	0.276	0.254	0.315	2.083	0.083	0.122	0.105	0.135	2.471	0.471	0.879	0.924	1.037
	DR (T.PS, F.OR)	2.008	0.008	0.374	0.576	0.576	1.997	0.003	0.182	0.304	0.303	2.035	0.035	0.980	1.175	1.175
	DR (F.PS, T.OR)	1.993	0.007	0.199	0.199	0.199	1.994	0.006	0.089	0.089	0.089	2.043	0.043	0.634	0.636	0.638
	DR (F.PS, F.OR)	2.553	0.553	0.348	0.315	0.640	2.543	0.543	0.153	0.136	0.564	2.514	0.514	1.113	1.357	1.450
	DR (GBM, T.OR)	1.997	0.003	0.194	0.204	0.203	1.999	0.001	0.087	0.088	0.088	2.022	0.022	0.608	0.619	0.618
	DR (GBM, F.OR)	2.236	0.236	0.275	0.249	0.345	2.114	0.114	0.122	0.104	0.156	2.435	0.435	0.897	0.946	1.040

Note: T.OR indicates true outcome regression; F.OR indicates false outcome regression.

Table 2.4: ATT estimates along with standard errors obtained using different propensity score based approaches for the Lalonde study using two methods of propensity score estimation (logistic regression and GBM).

	Logistic regression		GBM	
	Estimate	Std. Error	Estimate	Std. Error
Matching	635	747	608	761
Regression	1843	873	1967	1289
Strat	1377	831	563	1342
IPW	1274	739	628	941

Table 2.5: ATE estimates along with standard errors obtained using different propensity score based approaches for the Lindner data set using two methods of propensity score estimation (logistic regression and GBM).

	Logistic regression		GBM	
	Estimate	Std. Error	Estimate	Std. Error
Regression	944	848	562	884
Strat	1220	856	799	918
IPW	3628	1390	2040	999
IPWN	1285	1108	850	954
DR	1921	1085	862	910

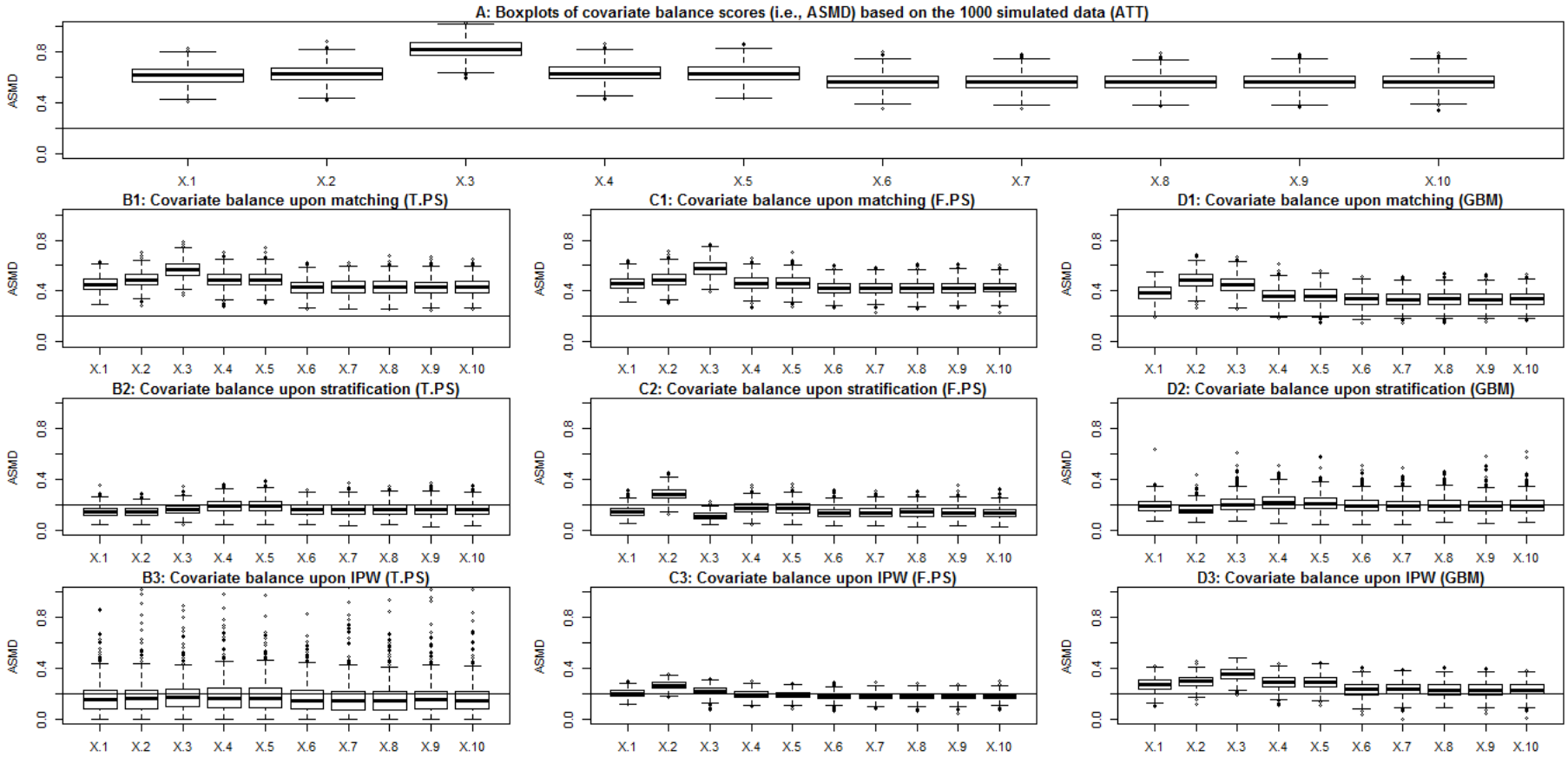

Figure 2.1: Assessment of covariates balancing of different methods for estimating ATT based on 1000 simulated data with a sample size of 1000, where the covariates are dependently normally distributed. Panel A is the boxplot of the balancing scores (i.e., ASMD) for each covariate; Panels B1-B3 are the boxplots of the balancing scores for each covariate upon matching, stratification, and IPW, respectively, when propensity score is estimated using the true logistic regression model; Panel C1-C3 are the boxplots balancing scores for each covariate when propensity score is estimated using the false logistic regression model; Panel D1-D3 are the boxplots of the balancing scores for each covariate when propensity score is estimated using the GBM model.

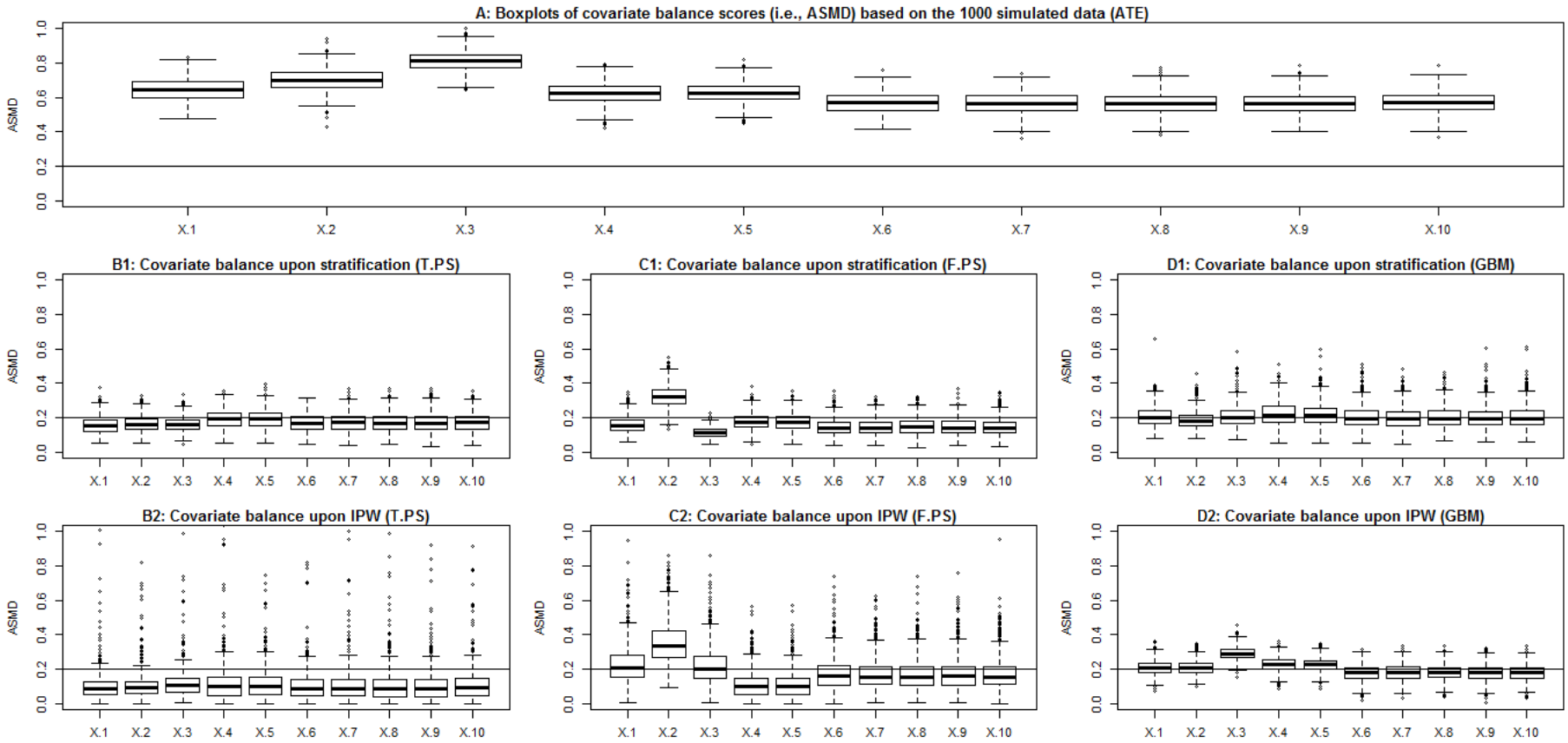

Figure 2.2: Assessment of covariates balancing of different methods for estimating ATE based on 1000 simulated data with a sample size of 1000, where the covariates are dependently normally distributed. Panel A is the boxplot of the balancing scores (i.e., ASMD) for each covariate; Panel B1-B2 are the boxplots of the balancing scores for each covariate upon stratification, and IPW, respectively, when propensity score is estimated using the true logistic regression model; Panel C1-C2 are the boxplots of the balancing scores for each covariate when propensity score is estimated using the false logistic regression model; Panel D1-D2 are the boxplots of the balancing scores for each covariate when propensity score is estimated using the GBM model.

CHAPTER 3
ESTIMATION OF AVERAGE TREATMENT EFFECTS AMONG MULTIPLE TREATMENT GROUPS BY USING ADAPTIVE ENSEMBLE METHOD

3.1 Introduction

Randomized control trials (RCT) are considered as the gold standard to determine the treatment effect between different treatment groups. In a RCT the subjects are randomly assigned to treatment groups and it is assumed that all confounding baseline covariates either measured or unmeasured are balanced (Austin, 2011), and therefore treatment effect can be directly estimated by the difference of observed group means. However, it is not always feasible to conduct an RCT due to ethics, cost, and patient preferences (Feng et al., 2012). On the other hand, observed data from patients under different treatments in a natural health care setting, which is termed as observational study, can be available. In a natural health care setting the treatment choice for each patient usually depends on the patient's characteristics and the doctor's preferences. Thus the assumption on covariate balance may not be valid any more, and how to assess treatment effect properly becomes challenging.

To assess treatment effects based on observational study, Rosenbaum and Rubin (1983) introduced a novel idea of propensity score to balance the covariates. The propensity score is the probability of treatment assignment conditional on the observed baseline covariates (Austin, 2011). Several methods based upon propensity score have been developed to assess treatment effect from observational

studies. These methods include matching (Rosenbaum and Rubin, 1983 and 1985), regression with propensity score as a covariate (Rosenbaum and Rubin, 1983; Rosenbaum, 1987), stratification (Rosenbaum and Rubin, 1983 and 1984; Lunceford and Davidian, 2004), inverse probability weighting (IPW) (Rosenbaum, 1987; Lunceford and Davidian, 2004), and the doubly robust (DR) method (Lunceford and Davidian, 2004). Rosenbaum and Rubin (1983) used logistic regression to estimate the propensity score. Using logistic regression may lead to a biased estimator of propensity score if the model is mispecified. Several different non-parametric techniques such as generalized boosting model (GBM), neural networks, classification and regression trees (CART), pruned CART, bootstrap aggregated (bagged) CART and random forests are used to improve the propensity score estimation (McCaffrey et al., 2004; Lee et al., 2010; Setoguchi et al., 2008).

The propensity score framework developed by Rosenbaum and Rubin (1983) was for assessing treatment effect between two treatment groups. Whereas, Imbens (2000) outlines the framework for estimating treatment effects via the generalized propensity score when there are multiple treatment groups. The generalized propensity score is the conditional probability of receiving a particular level of the treatment, given pre-treatment variables (Imbens, 2000). Imbens (2000) used the multinomial logistic regression to estimate the unknown generalized propensity score. On the other hand, McCaffrey et al. (2013) used GBM to improve the quality of the estimate of the unknown generalized propensity score. It has been seen that only multinomial logistic regression (Feng et al., 2012; Imbens, 2000) and GBM (McCaffrey et al., 2013) are used to estimate the generalized propensity score. In this project, an adaptive optimal ensemble method (Datta et al., 2010) is developed to estimate the generalized propensity score by ranking the performances of multinomial logistic regression, random forests, and GBM in balancing covariates. Based on the estimated generalized propensity score, we apply IPW, stratification

and doubly robust methods to estimate the average treatment effect (ATE). The ATE is defined as the mean of the individual causal effects in the whole population.

This chapter is structured as follows. In Section 3.2 we present the assumptions for causal inference for multiple treatment groups and develop the ensemble method for estimating the generalized propensity score. In Section 3.3, we present different methods for estimating ATE among multiple treatment groups. In Section 3.4, a simulation study is carried out to examine the performances of these methods. In Section 3.5, a case study is carried out to examine the treatment effects among three different treatments on patients with spinal fusion. Last section is devoted to a discussion.

3.2 Basic assumptions for causal inferences and an ensemble classifier for estimating generalized propensity scores.

In this section we first present the notations and basic assumptions for estimating the treatment effects. Then we develop an ensemble method to estimate the generalized propensity score. The ensemble method is obtained by ranking the overall performance of each different generalized propensity score estimating method.

3.2.1 Notation and assumptions

Imbens (2000) outlines the framework for estimating the treatment effect via the generalized propensity score when there are multiple treatment groups. The generalized propensity score is the conditional probability of receiving a particular level of the treatment given pre-treatment variables (Imbens, 2000), which can be written as:

$$r(t, X) = Pr(T = t|X). \tag{3.1}$$

Here the random variable T denotes the treatment received and t is the realization of T and X is the vector of pre-treatment covariates. If we let M denote all possible treatment choices, then $t \in \{1, \cdots, M\}$.

To estimate treatment effects, the following assumptions are required:

- Positivity (sufficient overlap): a subject has a non-zero probability of receiving each treatment. Mathematically, it can be written as:

$$0 < Pr(T = t|X) < 1, \quad t \in \{1, \cdots, M\}. \tag{3.2}$$

  Here $\sum_{t=1}^{M} Pr(T = t|X) = 1$.

- Strong ignorability of treatment assignment (SITA): there are no unmeasured or unknown confounders. In other words, the vector of covariates $\mathbf{X}$ includes all covariates which influence both the treatment selection and the potential outcome (McCaffrey et al., 2013).

3.2.2 An optimal ensemble method for estimating the generalized propensity score

The generalized propensity score plays the role of balancing covariates and dimension reduction. We first introduce three different methods (i.e., multinomial logistic regression, random forests, and GBM) for estimating the generalized propensity score, and methods to evaluate the balance of covariates. Then we develop the optimal ensemble method to estimate generalized propensity scores which is based on the ranking of the three different generalized propensity score estimating methods in terms of balance of covariates. In the following, let us denote (x_i, y_i, t_i) as the observations for the i^{th} subject, where $i = 1, \cdots, n$. Here, x_i is the vector of the p covariates for the i^{th} subject, y_i and t_i, respectively denote the observed outcome

and the observed treatment for the i^{th} subject.

3.2.2.1 Proposed methods for estimating the generalized propensity score.

(i) *Multinomial logistic regression for estimating the generalized propensity score*

Multinomial logistic regression is the most commonly used method to estimate the generalized propensity score. Assume that there are M possible treatment selections, and set $t = 1$ as the reference group. The multinomial logistic regression fits $M - 1$ equations (Feng et al., 2012):

$$ln(\frac{r(t,X)}{r(1,X)}) = ln\left(\frac{P(T=t|X)}{P(T=1|X)}\right) = \beta_{t0} + \beta_{t1}X, \quad \text{for} \quad t = 2, \cdots, M. \tag{3.3}$$

Here $\sum_{t=1}^{M} r(t,X) = \sum_{t=1}^{M} Pr(T=t|X) = 1$. One may estimate the parameters β_{t0} and β_{t1} $(t = 2, \cdots, M)$ by maximizing the likelihood function based on the observed covariates and treatment selection data. From equation (3.3), one can get

$$\hat{r}(t,X) = \hat{r}(1,X)e^{\hat{\beta}_{t0}+\hat{\beta}_{t1}X}. \tag{3.4}$$

From the constraint $\sum_{t=1}^{M} \hat{r}(t,X) = 1$, one can obtain

$$\hat{r}(1,X) = \frac{1}{1+\sum_{t=2}^{M} e^{\hat{\beta}_{t0}+\hat{\beta}_{t1}X}}, \tag{3.5}$$

and

$$\hat{r}(t,X) = \frac{e^{\hat{\beta}_{t0}+\hat{\beta}_{t1}X}}{1+\sum_{t=2}^{M} e^{\hat{\beta}_{t0}+\hat{\beta}_{t1}X}} \quad \text{for} \quad t = 2, \cdots, M. \tag{3.6}$$

(ii)*Random forests method for estimating the generalized propensity score*

Random forests method is a machine learning method that builds a large number of de-correlated trees then averages them to obtain the estimated generalized propensity scores (Hastie et al., 2009). Each tree in the random forests is built from a bootstrap sample of size n selected with replacement from the original data and m variables randomly selected from the p covariates in X. Here m is a small

number which is usually taken as $\sqrt{p}$ for classification (Hastie et al., 2009). Then, generalized propensity scores for an observation are estimated as the proportion of the subjects in the corresponding terminal node with a specified treatment. The process is repeated B times. The estimate of the generalized propensity score based on random forests is the average of these B estimates. We used the *randomForest* function available in the R package "randomForest" to estimate the generalized propensity score. The number of trees in the *randomForest* function was set as 5000 when estimating the generalized propensity score.

(iii) *Generalized boosted model for estimating the generalized propensity score*

The generalized boosted model (GBM) has been applied to estimate the propensity score for two and multiple treatment groups (McCaffrey et al., 2004; McCaffrey et al., 2013). GBM is an automated data-adaptive algorithm, which uses regression trees as weak predictors and capture nonlinear and interactive effect of the covariates (McCaffrey et al., 2004; Burgette et al., 2015). We used the *gbm* function avaiable in the R package gbm to construct the GBM model with an interaction depth of three and a shrinkage factor of 0.05. The fitted GBM model was used to estimate the generalized propensity score.

3.2.2.2 Assessing the balance of a covariate

The main objective of propensity score is to balance the covariates. To assess the covariate balance we used the absolute standardized mean difference (ASMD) and Kolmogrov-Smirnov statistic (KS) (McCaffrey et al., 2013). The ASMD statisitc for t^{th} treatment group and k^{th} covariate is given by

$$ASMD_{tk} = \frac{|\bar{X}_{kt} - \bar{X}_{kP}|}{\hat{\sigma}_{kP}}, \tag{3.7}$$

where $t = 1, \cdots, M; k = 1, \cdots, p$; $\bar{X}_{kt} = \sum_{i=1}^{n} I_{\{T_i=t\}} X_{ki} w(t; x_i) / \sum_{i=1}^{n} I_{\{T_i=t\}} w(t; x_i)$ is the weighted mean of the k^{th} covariate in t^{th} group with $w(t; x_i) = \frac{1}{\hat{r}(t,x_i)}$. Here

$\bar{X}_{kP}$ and $\hat{\sigma}_{kP}$ are the unweighted mean and standard deviation for the k^{th} covariate pooled across all treatment groups. The weight $w(t; x_i)$ for the i^{th} observation represents the number of subjects in the population who had the same covariates. In general, ASMD greater than 0.20 indicates that the covariate is unbalanced (McCaffrey et al., 2013).

Covariate balance can also be assessed by using the Kolmogorov-Smirnov (KS) statistic (McCaffrey et al., 2013). The KS statistic for k^{th} covariate and t^{th} treatment is given by

$$KS_{tk} = sup|EDF_{tk}(x) - EDF_{Pk}(x)|, \tag{3.8}$$

whereEDF_{tk} is the weighted empirical distribution function for k^{th} covariate in the t^{th} treatment group, which is given by $EDF_{tk}(x) = \sum_{i=1}^{n} w(t; x_i) I_{\{T_i=t\}} I_{(X_{ik} \leq x)} /$ $\sum_{i=1}^{n} w(t; x_i) I_{\{T_i=t\}}$, where $w(t; x_i) = 1/\hat{r}(t, x_i)$. $EDF_{Pk}(x)$ is the unweighted empirical ditribution function for the k^{th} covariate pooled across all treatment groups, which is given by $EDF_{Pk}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(X_{ik} \leq x)}$.

3.2.2.3 The optimal ensemble method

Recall that the generalized propensity score plays the role of balancing covariates. A covariate balance is examined by the absolute standardized mean difference (ASMD) and the Kolmogorov-Smirnov statistic. Multiple covariate balance metrics are used to evaluate the performances of the three different methods for estimating the generalized propensity score. It is challenging to obtain an optimal generalized propensity score estimating method with respect to all performance metrics. Datta et al. (2010) suggested an adaptive optimal ensemble classifier, which uses bootstrap aggregation (bagging) and rank aggregation (Datta et al., 2010; Pihur et al., 2009) to determine the optimal ensemble classifier in the presence of multiple performance metrics. To estimate the generalized propensity score using an optimal ensemble method we applied the algorithm stated below. Assume that we have J

methods and Q performance metrics. The algorithm can be implemented as follows:

Step 1. Generate 1000 realizations of $\boldsymbol{X}$, treatment assignment T and response variable Y. The detailed explanation of how to generate $\boldsymbol{X}, T \text{and} Y$ is outlined in section 3.4.

Step 2. Generate a bootstrap sample of size 1000 with replacement from the data generated in Step 1. It is important that each treatment group is represented in the bootstrap sample. Otherwise, sampling should be repeated to ensure that each treatment group is represented in the bootstrap sample. Note that in the bootstrap sample, approximately 2/3 of the sampling units from the original dataset are included in the bootstrap sample (Efron and Tibshirani, 1993). The rest of the 1/3 of the observations are left out of the bootstrap sample, which is known as the out-of-bag (OOB) sample (Efron and Tibshirani, 1993).

Step 3. Estimate the generalized propensity score for the bootstrap sample generated in Step 2 using J different generalized propensity score estimating methods. The methods include multinomial logistic regression, random forests, and GBM.

Step 4. Calculate the Q performance metrics for each method in Step 3 using the OOB sample. For our study the performance metrics are absolute standardized difference mean (ASMD) and Kolmogorov-Smirnov statistic (KS). Each peformance metric is calculated for the J different estimating methods. The performances of the J methods based on the performance metric can be ranked. Repeat this procedure for each performance metric, Q ordered lists of size J can be formed, say, $L_1, \cdots, L_Q$.

Step 5. The ordered lists determined in Step 4 are aggregated using the weighted rank aggregation method, which gives the best method for estimating the

generalized PS based on the b^{th} bootstrap sample, denoted by $A^b_{(1)}$. $A^b_{(1)}$ is the optimal method in the J ordered list, which minimizes the weighted rank aggregation quantity:

$$\Phi(\delta) = \sum_{i=1}^{Q} \widetilde{w}_i d(\delta, L_i). \tag{3.9}$$

Here δ is any valid ordered list of size J, d is a distance function, and $\widetilde{w}_i$ are the weights. The weights, $\widetilde{w}_i$ $(i = 1, \cdots, Q)$ provide a great flexibility in rank aggregation. In our simulation and case study, we set $\widetilde{w}_i = 1$.

Step 6. Repeat Step 2 to Step 5 B times to form a list of B best models to estimate the generalized propensity score, denoted by $A^1_{(1)}, \cdots, A^B_{(1)}$.

Step 7. The final method is selected by a majority vote.

Step 8. Estimate the generalized propensity score for the complete data set generated in Step 1 by using the method selected in Step 7.

Step 9. Using the complete data set generated in Step 1 and generalized propensity scores estimated in Step 8, estimate the ATE using inverse probability weighting (IPW), stratification, adn doubly robust (DR) methods. These generalized propensity score based methods are presented in the following section.

Repeat Step 1 to Step 9 1000 times. We reported the mean of 1000 estimates of ATE by each method.

3.3 Generalized propensity score based statistical methods for estimating ATE

The ATE of treatment t' relative to t'' is the comparison of mean outcomes had the entire population been observed under one treatment t' versus the other treatment t'' (McCaffrey et al., 2013). Mathematically it can be written as:

$$ATE_{t',t''} = \mu_{t',t''} = E(Y_{t'} - Y_{t''}) = E(Y_{t'}) - E(Y_{t''}) = \mu_{t'} - \mu_{t''}. \tag{3.10}$$

In the following, we present the commonly used generalized propensity score based methods for estimating ATE, which include stratification, IPW, and DR.

3.3.1 Stratification method for estimating ATE

Stratification could be used to estimate the pairwise ATE when there are M treatment groups. To estimate the μ_t for the t^{th} treatment group using stratification, the strata are first constructed based on the quantiles of the $r(t,x)$ (Yang et al., 2016). Thus, subjects with similar generalized propenisty scores are placed into one stratum. The mean of the observed outcomes for the treatment t is estimated in each stratum using

$$\hat{\mu}_{kt} = \frac{1}{N_{kt}} \sum_{k^{th} strata} I_{\{T_i=t\}} y_i, \tag{3.11}$$

where N_{kt} is the number of subjects in t^{th} treatment group and k^{th} stratum. The mean outcome of the entire population when treated with treatment t is estimated by

$$\hat{\mu}_{t,Strat} = \sum_{k=1}^{K} \frac{N_k}{N} \hat{\mu}_{kt}, \tag{3.12}$$

where N_k is the total number of subjects in the k^{th} strata and N is the total number of subjects in the entire sample. The ATE between treatments t' and t'' can be estimated as the difference between $\hat{\mu}_{t',Strata}$ and $\hat{\mu}_{t'',Strata}$.

3.3.2 Inverse probability weighting method for estimating ATE

McCaffrey et al. (2013) provided a pairwise estimate for ATE when there are multiple treatment groups. The mean outcome of the entire population when treated with treatment t (McCaffrey et al., 2013), say μ_t, can be estimated as

$$\hat{\mu}_t = \frac{\sum_{i=1}^{n} I_{\{T_i=t\}} y_i w(t,x_i)}{\sum_{i=1}^{n} I_{\{T_i=t\}} w(t,x_i)}, \tag{3.13}$$

where

$$w(t; x_i) = \frac{1}{r(t, x_i)}. \tag{3.14}$$

The ATE between treatments t' and t'' can be estimated as the difference between $\hat{\mu}_{t'}$ and $\hat{\mu}_{t''}$.

3.3.3 Doubly robust method for estimating ATE

DR estimator is an amendment to the IPW estimator (Robins et al., 1994). The DR estimator involves a regression model for the outcome variable. DR estimator remains consistent if either the generalized propensity score or the outcome regression is correctly specified. The DR estimator for the t^{th} treatment group is given by (Nilsson, 2013)

$$\hat{\mu}_{t,DR} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{I_{\{T_i=t\}} y_i}{\hat{r}(t, x_i)} - \frac{I_{\{T_i=t\}} - \hat{r}(t, x_i)}{\hat{r}(t, x_i)} m_t(x_i, \hat{\alpha}_t) \right), \tag{3.15}$$

Here $m_t(x_i, \hat{\alpha}_t)$ is the outcome regression model, outlined in Section 3.4, for the outcome variable Y on X for the treatment group. $\hat{r}(t, x_i)$ is the generalized propensity score for the i^{th} subject and t^{th} group. The ATE between treatments t' and t'' can be estimated by the difference between $\hat{\mu}_{t',DR}$ and $\hat{\mu}_{t'',DR}$.

3.4 Simulation study

3.4.1 Simulation scenario

We conducted a simulation study to examine the performance of different methods for estimating ATE. In this simulation study we considered 5 continuous and 1 binary covariates, say $\boldsymbol{X} = (X_1, X_2, \cdots, X_6)$. Here X_1, X_2, and X_3 follow a

multivariate normal with zero mean and covariance matrix

$$\begin{bmatrix} 2 & 1 & -1 \\ 1 & 1 & -0.5 \\ -1 & -0.5 & 1 \end{bmatrix};$$

X_4 follows a uniform distribution with the support on the interval $(-3, 3)$, that is, $X_4 \sim U(-3, 3)$; $X_5 = X_5^* - 1$, where X_5^* follows Chi square distribution with 1 degree of freedom; and $X_6 = X_6^* - 0.5$, where X_6^* follows a Bernoulli distibution with parameter 0.5. The setting used here similar to the settings used by Yang et al. (2016). The simulations are carried out as follows:

Step 1. Generate 1000 realizations of $\boldsymbol{X} = (X_1, X_2, \cdots, X_6)$.

Step 2. Using the 1000 realizations of $\boldsymbol{X}$ generated in Step 1, calculate treatment selection probabilities using the multinomial logistic regression model:

$$Pr(T = 1) = \frac{1}{1+exp(\widetilde{X}^T\beta^{(1)})+exp(\widetilde{X}^T\beta^{(2)})}, \tag{3.16}$$

$$Pr(T = 2) = \frac{exp(\widetilde{X}^T\beta^{(1)})}{1+exp(\widetilde{X}^T\beta^{(1)})+exp(\widetilde{X}^T\beta^{(2)})}, \tag{3.17}$$

$$Pr(T = 3) = \frac{exp(\widetilde{X}^T\beta^{(2)})}{1+exp(\widetilde{X}^T\beta^{(1)})+exp(\widetilde{X}^T\beta^{(2)})}. \tag{3.18}$$

Here $\widetilde{X} = (X_1, X_2, X_3, X_4, X_1^2, X_2^2, X_2X_3, X_3X_4X_5^*)^T$, $\beta^{(1)} = (0.25, 0.25, 0.25, -0.25, 0.25, 0.25, 0.25, 0.25)$, and $\beta^{(2)} = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$.

Step 3. Generate 1000 realizations for the treatment assignment T from multinomial distribution using the treatment selection probabilities calculated in Step 2.

Step 4. Generate 1000 realizations of the response variable based on the 1000 realizations of X generated in Step 1 and the 1000 realization of treatment assignments. Following outcome regressions models were used:

$$Y = \begin{cases} X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + \epsilon, & \text{if } T = 1; \\ 2X_1 + 3X_2 + X_3 + 2X_4 + 2X_5 + 2X_6 + \tau_1 + \epsilon, & \text{if } T = 2; \\ 3X_1 + X_2 + 2X_3 - X_4 - X_5 - X_6 + \tau_2 + \epsilon, & \text{if } T = 3. \end{cases} \tag{3.19}$$

Here ϵ follows a normal distribution with zero mean and variance 1, and (τ_1, τ_2) are known values. In order to assess the DR method, we purposely fitted following outcome regression models, which form the false outcome regression models:

$$Y = \begin{cases} X_1 + X_4 + \epsilon, & \text{if } T = 1; \\ X_2 + X_5 + \epsilon, & \text{if } T = 2; \\ X_3 + X_6 + \epsilon, & \text{if } T = 3. \end{cases} \tag{3.20}$$

Step 5. Estimate the generalized propensity score using multinomial logistic regression, random forests, GBM, and ensemble classifier, respectively.

Step 6. Estimate the ATE using the methods described in Section 3.3 and their standard errors using the bootstrap method.

Repeat Step 1 to Step 6 1000 times and for each specification of (τ_1, τ_2). (τ_1, τ_2) is taken as $(0, 0), (0, 0.5), (0.5, 0.5), \text{and}(0.5, 1)$ respectively, so that we can examine the size and the power of hypothesis test for $H_0 : \tau_1 = \tau_2 = 0$ against $H_a : \tau_1 \neq 0$ or $\tau_2 \neq 0$.

For each set of (τ_1, τ_2), we report the mean of 1000 estimates of ATE for each method (see Table 3.1), the absolute bias based on the 1000 estimates of ATE for each method (see Table 3.2), the boxplot of the 1000 estimated standard errors (see Figure 3.1) for the case (τ_1, τ_2) at $(0, 0)$, the true coverage rate of the estimated 95% confidence interval (see Table 3.3), and the root mean square error (RMSE) of the estimates for each method (see Table 3.4).

3.4.2 Simulation results

In this subsection, we discuss simulation results for the scenarios outlined in the previous subsection. The generalized propensity scores are estimated using the multinomial logistic regression, random forests, GBM, and the adaptive ensemble

method. Each generalized propensity score estimating method is combined with IPW, stratification, and DR method with correctly and incorrectly specified models (FOR, TOR). Thus there are 16 cases. Table 3.1 presents the mean of the 1000 estimated ATE under each specification of (τ_1, τ_2) for all 16 cases. Table 3.2 provides the absolute bias for the estimates of ATE. For IPW method, when the generalized propensity scores was estimated using the multinomial logistic regression the estimates between treatment 3 and treatment 1 were close to the true value, whereas ATE estimate for the other two pairs were biased across all specifications of (τ_1, τ_2). When the generalized propensity score was estimated using the random forests and GBM, the pairwise estimates for ATE using the IPW method was biased irrespective of the specification for (τ_1, τ_2). When the generalized propensity scores were estimated using the ensemble method, all the estimates were least biased across all specifications of (τ_1, τ_2).

For the stratification method, when the generalized propensity score was estimated using GBM, estimates were least biased and were close to the true value across all specifications of (τ_1, τ_2). For the DR method with false outcome regression model, when the generalized propensity scores were estimated using the ensemble classifier, the pairwise estimates of ATE were the least biased and were very close to the true value of ATE across all specifications of (τ_1, τ_2). For the DR method with true outcome regression model, the pairwise estimates of ATE were close to the underlying true value of ATE across all specifications of (τ_1, τ_2) irrespective of the generalized propensity scores estimating method.

The boxplots of the 1000 standard error for estimating ATE are presented in Figure 3.1 for each of the 16 combined methods. The standard error for each method for estimating ATE was obtained using the bootstrap method. Panel A1 presents the boxplots of the standard errors for the estimated treatment effect between treatment 2 and treatment 1, Panel A2 represents the boxplots of the standard error for the

estimated treatment effect between treatment 3 and treatment 1, and Panel A3 is for treatment 3 versus treatment 2. When the generalized propensity score is estimated using multinomial logistic regression, GBM, and an ensemble classifier the standard errors were small, whereas when generalized propensity score was estimated using the random forests, the standard errors were large.

Table 3.3 reports the true coverage rates of 95% confidence intervals based on 1000 simulated data sets. From Table 3.3, it is clear that the true coverage was far from the nominal coverage when random forests were used to estimate the generalized propensity scores, which are probably caused by large standard errors (see Figure 3.1). On the other hand, the true coverage rate is close to the nominal 95% coverage rate when the generalized propensity scores is estimated using the ensemble classifier. The RMSE is reported in Table 3.4. From Table 3.4, the RMSE is large when the generalized propensity score was estimated using random forests, whereas the RMSE were low and were close to each other when the generalized propensity score was estimated using multinomial logistic regression, GBM or the ensemble method.

3.5 Case study

A MarketScan insurance claim data set is used to examine the performance of the different methods for estimating ATE. MarketScan data set contains insurance claims for the spinal fusion for spinal degenerative disease. In the MarketScan data set the spinal fusion was done using three different treatments. These treatments include bone morphogenetic proteins (BMP), autograft, and allograft. The data set contains insurance claims made by Medicare, Medicaid, and commercial insurance companies over the years 2001 to 2011. The outcome variables was the total cost spent on the outpatient for the procedure. Each patient was treated with one of

the three treatments. The following covarites were used for the analysis: (i) age; (ii) calender year when surgery happened; (iii) gender, 1 for female and 0 for male; (iv) fusion, three different methods (i.e., interbody fusion, posterior fusion, and circumferencial fusion) indicated by two dummy variables with circumferencial fusion as the reference group; (v) type of insurance, which includes medicare, medicaid, and commercial; and (vi) the Charlson comorbidity score.

There were 54105 subjects in the MarketScan dataset who had a spinal degenerative disease, and was treated with only one of the three treatments: BMP, autograft, and allograft. The observations for the 54105 patients forms the data set for comparisons between different treatments. Among these 54105 subjects 32182 were females and 21923 were males and, 9929 insurance claims were from Medicare, 4485 insurance claims were from Medicaid, and 39691 insurance claims were from commercial insurance companies. Also among these 54105 subjects, 11094 were treated with BMP, 18690 were treated with autograft and 24321 subjects were treated with allograft.

The comparisons of the three treatments based on the MarketScan data using different statistical methods are reported in Table 3.5. Based on the simulation studies in Section 3.4, when the generalized propensity scores are estimated using the ensemble method, ATE was the least biased using the IPW or DR method. Whereas, when the generalized propensity score was estimated using GBM the estimates of the ATE was the least biased using stratification. From our simulation studies it is clear that the estimates of ATE were biased when random forests were used to estimate the generalized propensity scores.When the generalized propensity score was estimated using an ensemble classifer, the ATE estimates from IPW and DR were similar. When the generalized propensity score was estimated using GBM, the ATE estimates based on stratification were close to those based on the ensemble method IPW and DR. We drew conclusions based on the ATE estimates from

the ensemble method based on IPW and DR and the GBM based stratification. We concluded that the average difference of cost for an outpatient treated subject with BMP and autograft is roughly between 93 to 160 dollars, with standard error between 177-190 dollars, a nonsignificant difference. Similarly the average difference of cost for an outpatient treated sybject with BMP and allograft is roughly betweeen 1000 to 1068 dollars, with a standard error between 162.11-162.58 dollars, a significant difference. Whereas the average difference of cost spend for an outpatient treated subject with allograft and autograft is approximately between 840 to 957 dollars with a standard error between 152-168 dollars, a significant difference.

3.6 Discussion

In this article we propose an ensemble method to estimate the generalized propensity scores and investigate its performace by comparing with the existing methods. Based on our simulations and case studies, the estimates of ATE may vary from one method to another. When the generalized propensity score is estimated using random forests the estimates for ATE were usually less reliable. When the generalized propensity score was estimated using the ensemble method, the ATE estimates based on IPW or DR were the least biased. When the generalized propensity score was estimated using GBM, the ATE estimate based on the stratification was the least biased on comparing with IPW and DR. An ensemble method does provide promising results, although the computation is extensive.

Table 3.1: The average of the 1000 estimated ATE under each setting for (τ_1, τ_2). The generalized propensity score is estimated using the multinomial logistic regression, random forests, GBM and an adaptive ensemble method. Each generalized propensity score estimating method is combined with IPW, stratification, and DR method.

			Multinomial (FPS)				Random Forest				GBM				Ensemble Method			
(τ_1, τ_2)	Treatment Effect	True	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)
(0, 0)	2 vs 1	0	- 0.20	-0.11	-0.01	0.02	-0.19	0.11	-0.67	-0.02	-0.15	0.03	-0.65	0.02	-0.20	-0.11	-0.01	-0.00
	3 vs 1	0	-0.04	-0.05	-0.10	-0.04	-0.56	-0.12	-0.58	0.02	0.58	0.01	-0.62	0.01	-0.03	0.04	-0.09	-0.01
	3 vs 2	0	0.16	-0.06	-0.10	-0.02	-0.37	-0.24	0.09	-0.01	-0.43	0.03	0.04	-0.01	0.17	-0.06	-0.07	-0.01
(0, 0.5)	2 vs 1	0	-0.20	-0.10	0.05	-0.02	-0.19	0.12	-0.67	-0.02	-0.11	0.04	0.33	0.05	-0.20	-0.11	-0.03	-0.00
	3 vs 1	0.5	0.44	0.44	0.39	0.46	-0.05	0.08	-0.08	0.48	0.23	0.55	0.68	0.48	0.46	0.45	0.40	0.49
	3 vs 2	0.5	0.67	0.57	0.41	0.48	0.13	- 0.04	0.59	0.49	0.33	0.47	0.59	0.51	0.65	0.56	0.43	0.49
(0.5, 0.5)	2vs 1	0.5	0.27	0.40	0.49	0.48	0.33	0.42	-0.19	0.49	0.59	0.48	0.65	0.48	0.30	0.38	0.47	0.50
	3 vs 1	0.5	0.43	0.45	0.39	0.46	-0.12	0.06	-0.15	0.50	0.43	0.52	0.61	0.49	0.45	0.44	0.40	0.49
	3 vs 2	0	0.17	0.07	-0.08	-0.02	-0.45	-0.36	0.05	0.01	0.09	0.04	0.19	0.03	0.15	0.06	-0.07	-0.01
(0.5, 1)	2 vs 1	0.5	0.30	0.36	0.47	0.48	0.31	0.42	-0.17	0.48	0.60	0.47	0.67	0.47	0.30	0.38	0.47	0.50
	3 vs 1	1	0.93	0.94	0.90	0.96	0.44	0.28	0.42	0.98	1.10	0.98	0.89	1.02	0.95	0.94	0.90	0.99
	3 vs 2	0.5	0.67	0.57	0.40	0.48	0.13	-0.14	0.59	0.49	0.42	0.48	0.59	0.47	0.65	0.56	0.42	0.49

Note: FPS indicates false multinomial logistic regression; TOR indicates true outcome regression; FOR indicates false outcome regression.

Table 3.2: Absolute bias of the mean of 1000 estimates for ATE, where the generalized propensity score is estimated using the multinomial logistic regression, random forests, GBM and the adaptive ensemble method.

			Multinomial (FPS)				Random Forest				GBM				Ensemble Method			
(τ_1, τ_2)	Treatment Effect	True	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)
(0, 0)	2 vs 1	0	0.20	0.11	0.01	0.02	0.19	0.11	0.67	0.02	0.15	0.03	0.65	0.02	0.20	0.11	0.01	0.00
	3 vs 1	0	0.04	0.05	0.10	0.04	0.56	0.12	0.58	0.02	0.58	0.01	0.62	0.01	0.03	0.04	0.09	0.01
	3 vs 2	0	0.16	0.06	0.10	0.02	0.37	0.24	0.09	0.01	0.43	0.03	0.04	0.01	0.17	0.06	0.07	0.01
(0, 0.5)	2 vs 1	0	0.20	0.10	0.05	0.02	0.19	0.12	0.67	0.02	0.11	0.04	0.33	0.05	0.20	0.11	0.03	0.00
	3 vs 1	0.5	0.06	0.06	0.11	0.04	0.55	0.42	0.58	0.02	0.27	0.05	0.18	0.02	0.04	0.05	0.10	0.01
	3 vs 2	0.5	0.17	0.07	0.09	0.02	0.37	0.54	0.09	0.01	0.17	0.03	0.09	0.01	0.15	0.06	0.07	0.01
(0.5, 0.5)	2vs 1	0.5	0.23	0.10	0.01	0.02	0.17	0.08	0.69	0.01	0.09	0.02	0.15	0.02	0.20	0.12	0.03	0.00
	3 vs 1	0.5	0.07	0.05	0.11	0.04	0.62	0.44	0.65	0.00	0.07	0.02	0.11	0.01	0.05	0.06	0.10	0.01
	3 vs 2	0	0.17	0.07	0.08	0.02	0.45	0.36	0.05	0.01	0.09	0.04	0.19	0.03	0.15	0.06	0.07	0.01
(0.5, 1)	2 vs 1	0.5	0.20	0.14	0.03	0.02	0.19	0.08	0.67	0.02	0.10	0.03	0.17	0.03	0.20	0.12	0.03	0.00
	3 vs 1	1	0.07	0.06	0.10	0.04	0.56	0.72	0.58	0.02	0.10	0.02	0.11	0.02	0.05	0.06	0.10	0.01
	3 vs 2	0.5	0.17	0.07	0.10	0.02	0.37	0.64	0.09	0.01	0.08	0.02	0.09	0.03	0.15	0.06	0.08	0.01

Note: FPS indicates false multinomial logistic regression; TOR indicates true outcome regression; FOR indicates false outcome regression.

Table 3.3: True coverage rate based on the 1000 estimated 95% confidence intervals, where the generalized propensity score is estimated using the multinomial logistic regression, random forests, GBM and the adaptive ensemble method.

			Multinomial (FPS)				Random Forest				GBM				Ensemble Method			
(τ_1, τ_2)	Treatment Effect	True	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)
(0, 0)	2 vs 1	0	0.88	0.94	0.93	0.90	0.99	0.90	0.99	1	0.98	0.73	0.87	0.94	0.81	0.91	0.94	0.94
	3 vs 1	0	0.99	1	0.94	0.99	0.91	0.99	1	1	0.91	0.84	0.84	0.96	0.89	0.95	0.87	0.92
	3 vs 2	0	0.96	0.98	0.92	0.90	0.98	0.92	1	1	0.99	0.93	0.99	0.97	0.90	0.94	0.90	0.97
(0, 0.5)	2 vs 1	0	0.87	0.97	0.97	0.95	0.99	1	0.99	1	0.98	0.74	0.87	0.94	0.81	0.91	0.93	0.94
	3 vs 1	0.5	0.99	1	0.97	0.99	0.91	0.38	1	1	0.90	0.84	0.84	0.97	0.85	0.94	0.87	0.92
	3 vs 2	0.5	0.34	0.59	0.71	0.50	0.98	0.54	1	1	0.99	0.93	0.99	0.97	0.89	0.92	0.91	0.97
(0.5, 0.5)	2vs 1	0.5	0.85	0.93	0.95	0.95	1	1	1	1	0.98	0.75	0.87	0.94	0.81	0.91	0.93	0.95
	3 vs 1	0.5	0.98	0.75	0.87	0.94	0.99	0.99	0.99	1	0.90	0.84	0.84	0.97	0.88	0.94	0.87	0.93
	3 vs 2	0	0.91	0.98	0.95	0.96	0.98	0.82	0.99	1	0.99	0.97	0.99	0.97	0.89	0.99	0.91	0.97
(0.5, 1)	2 vs 1	0.5	0.98	0.75	0.87	0.94	0.99	0.99	0.99	1	0.98	0.75	0.87	0.94	0.80	0.91	0.93	0.94
	3 vs 1	1	0.90	0.84	0.84	0.97	0.90	0.02	1	1	0.90	0.84	0.84	0.97	0.88	0.93	0.87	0.92
	3 vs 2	0.5	0.99	0.93	0.99	0.97	0.98	0.38	1	1	0.99	0.97	0.99	0.97	0.89	0.98	0.91	0.97

Note: FPS indicates false multinomial logistic regression; TOR indicates true outcome regression; FOR indicates false outcome regression.

Table 3.4: Root mean square error of the 1000 estimates for each method, where the generalized propensity score is estimated using the multinomial logistic regression, random forests, GBM and the adaptive ensemble method.

			Multinomial (FPS)				Random Forest				GBM				Ensemble Method			
(τ_1, τ_2)	Treatment Effect	True	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)	IPW	STRAT	DR (FOR)	DR (TOR)
(0, 0)	2 vs 1	0	0.23	0.19	0.17	0.16	0.39	0.21	0.70	0.15	0.22	0.34	0.28	0.15	0.26	0.20	0.17	0.15
	3 vs 1	0	0.19	0.19	0.21	0.16	0.62	0.18	0.64	0.16	0.30	0.44	0.35	0.18	0.24	0.22	0.26	0.18
	3 vs 2	0	0.26	0.25	0.29	0.24	0.52	0.30	0.34	0.20	0.30	0.37	0.26	0.23	0.30	0.26	0.26	0.22
(0, 0.5)	2 vs 1	0	0.23	0.18	0.14	0.14	0.39	0.21	0.70	0.15	0.22	0.34	0.28	0.15	0.25	0.20	0.19	0.15
	3 vs 1	0.5	0.19	0.20	0.20	0.16	0.62	0.44	0.63	0.16	0.30	0.44	0.35	0.18	0.26	0.24	0.28	0.18
	3 vs 2	0.5	0.68	0.58	0.45	0.52	0.52	0.57	0.34	0.20	0.30	0.37	0.26	0.23	0.32	0.27	0.26	0.23
(0.5, 0.5)	2vs 1	0.5	0.26	0.21	0.16	0.15	0.39	0.21	0.70	0.15	0.22	0.34	0.28	0.15	0.25	0.20	0.19	0.15
	3 vs 1	0.5	0.23	0.23	0.25	0.18	0.62	0.44	0.63	0.16	0.30	0.44	0.35	0.18	0.26	0.24	0.28	0.18
	3 vs 2	0	0.31	0.26	0.27	0.23	0.52	0.57	0.34	0.20	0.30	0.37	0.26	0.23	0.32	0.27	0.26	0.23
(0.5, 1)	2 vs 1	0.5	0.22	0.35	0.27	0.14	0.39	0.18	0.70	0.15	0.22	0.34	0.28	0.15	0.25	0.20	0.19	0.15
	3 vs 1	1	0.30	0.44	0.34	0.18	0.62	0.73	0.64	0.16	0.30	0.44	0.35	0.18	0.26	0.24	0.28	0.18
	3 vs 2	0.5	0.30	0.36	0.27	0.23	0.53	0.67	0.34	0.20	0.30	0.37	0.26	0.23	0.32	0.27	0.26	0.23

Note: FPS indicates false multinomial logistic regression; TOR indicates true outcome regression; FOR indicates false outcome regression.

Table 3.5: ATE estimates (standard error) for the MarketScan data where the generalized propensity score is estimated using the multinomial logistic regression, random forests, GBM and the adaptive ensemble method.

	Multinomial			Random Forest			GBM			Ensemble Method		
Treatment Effect	IPW	STRAT	DR	IPW	STRAT	DR	IPW	STRAT	DR	IPW	STRAT	DR
Autograft vs BMP	159.75 (190.36)	205.53 (182.79)	147.60 (188.48)	1009.19 (881.57)	381.91 (293.64)	4190.27 (6575.20)	-16.33 (154.77)	93.31 (176.98)	148.50 (171.86)	159.75 (189.45)	205.53 (187.12)	147.60 (190.27)
Allograft vs BMP	999.33 (151.66)	1151.40 (155.80)	1066.53 (152.76)	1366.55 (751.26)	1240.83 (296.79)	7055.42 (5870.69)	902.82 (151.59)	1068.45 (162.13)	1046.49 (160.39)	999.33 (162.76)	1151.40 (160.76)	1066.53 (162.11)
Allograft vs Autograft	839.58 (161.90)	945.86 (157.34)	918.93 (161.31)	1009.19 (1020.07)	859.93 (267.87)	2865.16 (2765.48)	919.16 (137.93)	957.14 (152.88)	897.99 (152.54)	839.58 (168.23)	945.86 (163.28)	918.93 (159.23)

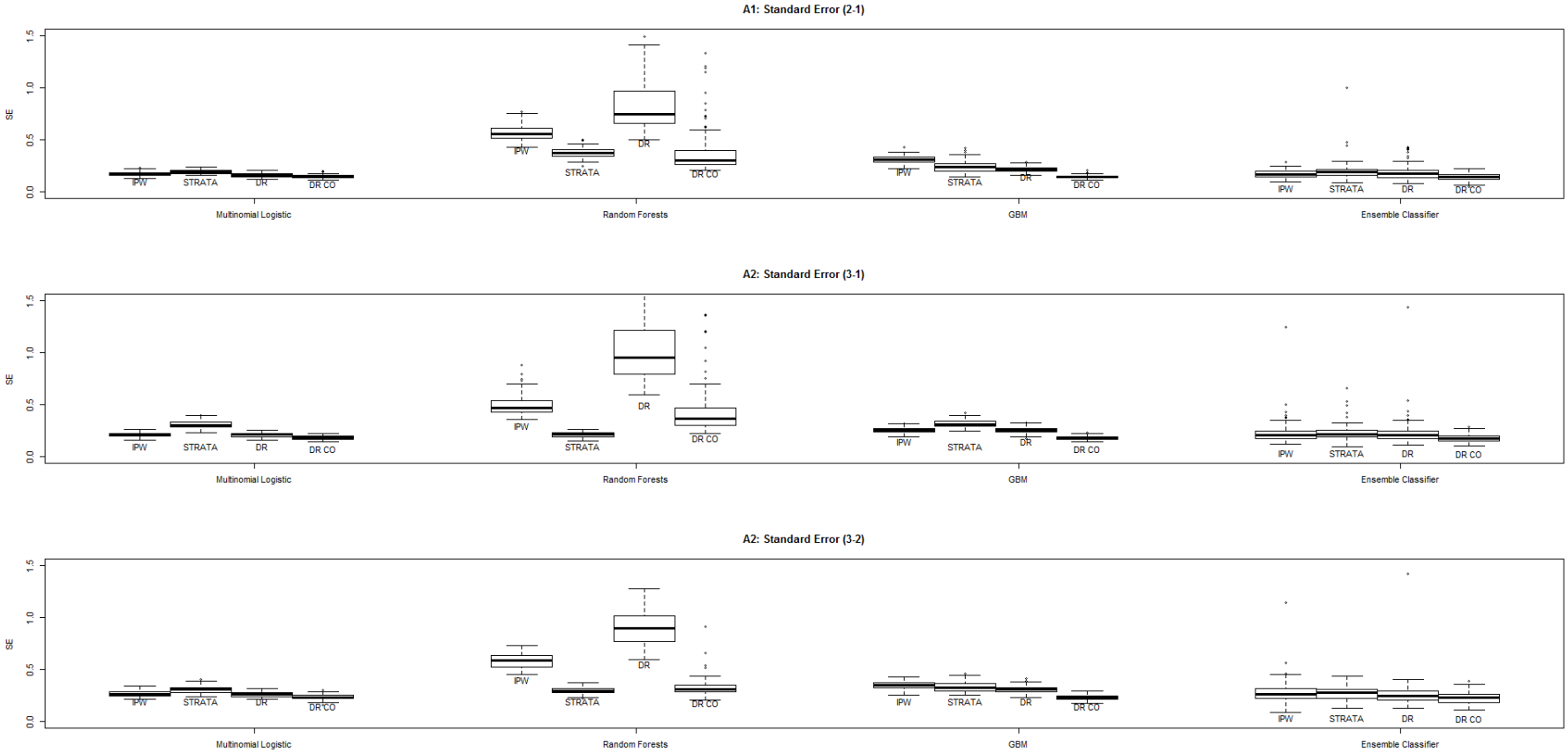

Figure 3.1: Standard error for estimating ATE based on 1000 simulated data with a sample size of 1000 with (τ_1, τ_2) at (0, 0). Panel A1 is the boxplot of the Standard errors when ATE is estimated between treatment 2 and treatment 1. Panel A2 is the boxplot of the standard errors when ATE is estimated between treatment 3 and treatment 1. Panel A3 is the boxplot of the standard errors when ATE is estimated between treatment 3 and treatment 2.

CHAPTER 4
CONCLUSION AND FUTURE RESEARCH

4.1 Conslusion

This dissertation consisted of two interconnected research projects. The first project was a study of propensity scores based statistical methods for estimating the ATE and ATT when there are two treatment groups. The propensity score was first estimated using the logistic regression or GBM. After estimating the propensity score the treatment effects were estimated using matching, propensity score adjusted regression, stratification, IPW and DR method. We conducted extensive simulation studies to determine the appropriate propensity score based method for estimating the treatment effect.

Rosenbaum and Rubin (1983) proposed the novel idea of propensity score to balance covariates and estimate treatment effects based on observational data. The propensity scores are unknown and are generally estimated using the logistic regression (Rosenbaum and Rubin, 1983). Since logistic regression is a parameteric model, misspecification of the parameteric model may lead to biased estimates of the propensity scores and treatment effects. Recently, non-parameteric machine learning methods such as GBM have been proposed to estimate the propensity score. The GBM uses an automated data-adaptive alogorithm, and selects the important variables and their interaction terms.

In Chapter 2, we examined different statistical methods to estimate the treat-

ment effects. Based on our simulations, when the propensity scores are estimated using GBM, IPW estimates for ATT are close to the true value of ATT. When the propensity score was estimated using GBM, ATE estimates from stratification, IPWN, and DR methods are close to the true value of ATE. From Chapter 2, we concluded that the stratification, IPWN, and DR provide reasonable estimates for ATE when propensity score are estimated using GBM. Whereas, the combination of GBM and IPW gives the appropriate estimates for ATT.

Chapter 3 was an extension of Chapter 2. In chapter 3, we extended our study to multiple treatment groups. We estimated the generalized propensity score using the multinomial logistic regression, random forests, GBM and an optimal ensemble method. To assess the balance among covariates, ASMD and KS statistic were calculated. After estimating the generalized propensity scores, ATE was estimated using the stratification, IPW, and DR method. Based on our simulation studies, the estimates of ATE may vary from one method to another. When the generalized propensity score is estimated using random forests, the estimates of ATE are usually less reliable. When the generalized propensity score was estimated using the ensemble method, the ATE estimates based on IPW and DR were the least biased. When the generalized propensity score was estimated using GBM, the ATE estimates based on the stratification was the least biased when comparing with IPW and DR.

4.2 Future research

Propensity score based methods are increasingly used to estimate the treatment effects observational studies. In chapter 2 we carried out the comparative study for estimating the treatment effects (ATT and ATE) using matching, propensity score adjusted regression method, stratification, IPW, and DR, where the

propensity score is estimated using logistic regression and GBM. Lee et al. (2010) also compared the machine learning methods such as CART, pruned CART, bootstrap aggregated CART, random forests and GBM to improve the estimation of propensity score. However, Lee et al. (2010) used only the IPW method to estimate the treatment effect (ATE). It would be interesting to compare other propensity score based methods such as regression, stratification, IPW, and DR, when the propensity score is estimate by the machine learning methods and the ensemble method developed in chapter 3. In our studies, we have considered continuous response variable. It will be interesting to examine the performance of different methods when the outcome variable is categorical variable such as count data or binary data.

In chapter 3 we proposed an optimal ensemble method to estimate the generalized propensity score. Some improvement of the algorithm could be carried out. For example, for each bootstrap sample in the ensemble algorithm, we could obtain the optimal generalized propensity score estimating method, estimate the generalized propensity score for each subject based on the optimal method, and then average the generalized propensity scores over the B bootstraps samples. I plan to carry out the simulations under the revised algorithm and anticipate the revised ensemble method will have a promising performance.

REFERENCES

Angrist, J. D. and Pischke, J. S. (2009). *Mostly harmless econometrics.* New Jersey, Princeton.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multlivariate Behavioral Research* **46**, 399-424.

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine* **33**, 1057-1068.

Datta, S., Pihur, V., and Datta, S. (2010). An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data. *BMC Bioinformatics* **11**.

Feng, P., Zhou, X. H., Zou, Q. M., Fan, M. Y. and Li, X. S. (2012). Generalized propensity score for estimating the average treatment effect of multiple treatments. *Statistics in Medicine* **31**, 681-697.

Gu, X. S. and Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: structure, distance and algorithm. *Journal of Computation and Graphical Statistics* **2**, 405-420.

Harder, V. S., Stuart, E. A. and Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods* **15**, 234-249.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning. 2^{nd} edition. NewYork: Springer.

Imai, K. and Van Dyk, D. A. (2004). Causal inference with general treatment regimes: generalizing the propensity score. *Journal of the American Statistical Association* **99**, 854-866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* **87**, 706-710.

Imbens, G. W. (2004). Nonparameteric estimation of average treatment effects under exogeneity : a review. *The Review of Economics and Statistics* **86**, 4-29.

Lalonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* **76**, 604-620.

Leacy, F. P. and Stuart, E. A. (2014). On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Statistics in Medicine* **33**, 3488-3508.

Lechner, M. (2001). Identification and estimation of causals effects of multiple treatments under the conditional independence assumptions. In

Econometric Evaluation of Labour Market Policies. Heidelberg: Physica-Verlag, 1-18.

Lee, B.K., Lessler, J., and Stuart, E.A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine* **29**: 337-346.

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effect: a comparative study. *Statistics in Medicine* **23**, 2937-2960.

McCaffrey, D. F., Ridgeway, G. and Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods* **9**, 403-425.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R. and Burgette, L. F. (2013) A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine* **32**, 3388-3414.

Nilsson, M. (2013). Causal inference in a 2^2 factorial design using generalized propensity score. Master Thesis, Uppsala University, Sweden.

Pan, W. and Bai, H. (2015). *Propensity score analysis: fundamentals and developments*. New York, Guilford.

Pihur,V., Datta, S., and Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation. *BMC Bioinformatics* **10**.

Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L. and Griffin, B. A. (accessed 12 March 2015) Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the twang package, https://cran.r-project.org/web/packages/twang /vignettes/twang.pdf.

Robins, J. M., Rotnitzky, A. and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of American Statistical Association* **47**, 663-685.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *The Journal of the American Statistician* **82**, 387-394.

Rosenbaum, P. R. (1989). Optimal matching for observational studies.*Journal of American Statistical Association* **84**, 10241032.

Rosenbaum, P. R. (1998). Propensity score. In: Armitage P and Colton T, *Encyclopedia of Biostatistics* New York, Wiley, 3551-3555.

Rosenbaum, P. R. (2009). *Design of observational studies.* New York, Springer Verlag.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects.*Biometrika* **70**, 41-55.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score.*Journal of the American*

Statistical Association **79**, 516524.

Rosenbaum, P. R. and Rubin, D. B.(1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* **39**, 33-38.

Setoguchi, S., Schneeweiss, S., Brookhart, M.A., Glynn, R.J., and Cook, E.F. (2008). Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and Drug Safety* **17**: 546-555.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science* **25**, 121.

Westreich, D., Lessler, J. and Funk, M. J. (2010). Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology* **63**, 826-833.

Williamson, E., Morley, R., Lucas, A. and Carpenter, J. (2011). Propensity score: From nave enthusiasm to intuitive understanding. *Statistical Methods in Medical Research* **21**, 273-293.

Yang, S., Imbens, G. W., Cui, Z., Faries, D. E. and Kadziola, Z. (2016). Propensity score matching and subclassification in observational studies with multi-level treatments. *Biometrics*, 1-11. doi: 10.1111/biom.12505

Zanutto, E., Lu, B., and Hornik, R. (2005) Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *Journal of Educational and Behavioral Statistics* **30**: 59-73.

APPENDIX

Table A1: Simulation results for estimating ATT, using the settings proposed by Leacy and Stuart, PS is estimated using the true logistic regression (T.PS),false logistic regression (F.PS), and GBM. The underlying ATT is -0.40.

		Sample size = 1000					Sample size = 5000				
Estimator	Method	Estimates	Bias	Std. Error	ESE	RMSE	Estimate	Bias	Std. Error	ESE	RMSE
ATT	Matching(T.PS)	-0.533	0.133	0.116	0.128	0.184	-0.490	0.090	0.032	0.041	0.099
	Matching(F.PS)	-0.531	0.131	0.115	0.116	0.175	-0.475	0.075	0.033	0.038	0.084
	Matching (GBM)	-0.532	0.132	0.074	0.081	0.155	-0.533	0.133	0.029	0.039	0.139
	Regression, $\overline{e(X_T)}$ (T.PS)	-0.309	0.091	0.135	0.107	0.140	-0.318	0.081	0.059	0.047	0.094
	Regression, $\overline{e(X_T)}$ (F.PS)	-0.400	0.000	0.124	0.107	0.083	-0.406	0.006	0.055	0.037	0.038
	Regression, $\overline{e(X_T)}$ (GBM)	-0.088	0.311	0.211	0.216	0.379	-0.219	0.180	0.077	0.064	0.191
	Stratification (T.PS)	-0.401	0.001	0.139	0.119	0.118	-0.414	0.014	0.060	0.052	0.054
	Stratification (F.PS)	-0.500	0.100	0.126	0.084	0.130	-0.508	0.108	0.055	0.039	0.115
	Stratification (GBM)	-0.423	0.023	0.330	0.246	0.247	-0.442	0.042	0.088	0.059	0.073
	IPW (T.PS)	-0.323	0.076	0.174	0.180	0.195	-0.351	0.048	0.077	0.089	0.101
	IPW (F.PS)	-0.527	0.127	0.111	0.107	0.166	-0.540	0.140	0.048	0.045	0.147
	IPW (GBM)	-0.517	0.117	0.080	0.080	0.142	-0.492	0.092	0.036	0.038	0.100

Note: Here T.PS is $logit(p_{treat}) = \gamma_0+\gamma_1 X_{.1}+\gamma_2 X_{.2}+\gamma_3 X_{.3}+\gamma_4 X_{.4}+\gamma_5 X_{.5}+\gamma_6 X_{.6}+\gamma_7 X_{.7}+\gamma_8 X_{.2}^2+\gamma_9 X_{.4}^2+\gamma_{10} X_{.7}^2+\gamma_{11} X_{.1}X_{.3}+\gamma_{12} X_{.2}X_{.4}+\gamma_{13} X_{.3}X_{.5}+\gamma_{14} X_{.4}X_{.6}+\gamma_{15} X_{.5}X_{.7}+\gamma_{16} X_{.1}X_{.6}+\gamma_{17} X_{.2}X_{.3}+\gamma_{18} X_{.3}X_{.4}+\gamma_{19} X_{.4}X_{.5}+\gamma_{20} X_{.5}X_{.6}$;

F.PS is $logit(p_{treat}) = \gamma_0 + \sum_{j=1}^{10} +\gamma_j X_{.j}$;

T.OR $Y = \eta_0 + \eta_1 X_{.1} + \eta_2 X_{.2} + \eta_3 X_{.3} + \eta_4 X_{.4} + \eta_5 X_{.8} + \eta_6 X_{.9} + \eta_7 X_{.10} + \eta_2 X_{.2}^2 + \eta_4 X_{.4}^2 + \eta_7 X_{.10}^2 + 0.5 \times \eta_1 X_{.1} X_{.3} + 0.7 \times \eta_2 X_{.2} X_{.4} + 0.5 \times \eta_3 X_{.3} X_{.8} + 0.7 \times \eta_4 X_{.4} X_{.9} + 0.5 \times \eta_5 X_{.8} X_{.10} + 0.5 \times \eta_1 X_{.1} X_{.9} + 0.7 \times \eta_2 X_{.2} X_{.3} + 0.5 \times \eta_3 X_{.3} X_{.4} + 0.5 \times \eta_4 X_{.4} X_{.8} + 0.5 \times \eta_5 X_{.8} X_{.9} - 0.4 \times Z + \epsilon$

Model Coefficient Values for T.OR:

$(\eta_0, \eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7) = (-1.386, 0.3, -0.36, -0.73, -0.2, 0.71, -0.19, 0.26)$

R Code for Chapter 2

```
#########################################################################
#########################################################################
## R Code producing results for chapter 2 when the propensity score
##is estimated using the logistic regression and variables are
## independently normally distributed with Zero mean and variance 1.
#########################################################################
#########################################################################


#########################################################################
#########################################################################
## Commands needed to run R Code on the Parallel computing environment
#########################################################################
#########################################################################

##args <- commandArgs()
## start w/args[3]
##Run <- as.numeric(args[3])
##Sim <- as.numeric(args[4])
##name <- paste("RESULT", Run,Sim, "RData", sep = ".")

#########################################################################
#########################################################################
## Packages needed to run the code
#########################################################################
#########################################################################

library(lme4)
library(Matrix)
library(Rcpp)
library(mvtnorm)
require(stats)

#########################################################################
#########################################################################
## Function to estimate standard error via bootstrap for IPW Methods
#########################################################################
#########################################################################


Var.bootstrap<- function(dataset,  nB=200)
{ ATT.ipw.ic.bs<- ATT.ipw.n.ic.bs<-ATT.ipw.c.bs<-ATT.ipw.n.c.bs<-  rep(NA,nB)
ATE.ipw.n.ic.bs<-ATE.ipw.ic.bs<- ATE.ipw.n.c.bs<-ATE.ipw.c.bs<-ATEDRCOR.bs<-rep(NA,nB)
  N<-nrow(dataset)
  for(b in 1:nB){
    bs.data<-as.data.frame(dataset[sample(nrow(dataset),N, replace=TRUE),])


#########################################################################
## Estimating the incorrect PS
#########################################################################
    ps.logit1<-glm(treat~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10, family=binomial, data = bs.data)
    bs.data$psic<- fitted.values(ps.logit1)
#########################################################################
## Estimating the correct PS
#########################################################################
    ps.logit1c <- glm(treat~x1+x2+x3+x1.sq+x2.sq+x.23+x.345, data = bs.data, family = binomial)
    bs.data$ps1<- fitted.values(ps.logit1c)
#################################################################
##  Estimating ATT using the IPW with incorrect PS
#################################################################
  temp.odd1<-(1-bs.data$treat)*bs.data$psic/(1-bs.data$psic)
  ATT.ipw.ic.bs[b]<-mean(bs.data$y[bs.data$treat==1])- mean(bs.data$y*temp.odd1)
  ATT.ipw.n.ic.bs[b]<-mean(bs.data$y[bs.data$treat==1])- sum(bs.data$y*temp.odd1)/sum(temp.odd1)

#################################################################
##  Estimating ATT using the IPWT with correct PS
#################################################################
  temp.oddc1<-(1-bs.data$treat)*bs.data$ps1/(1-bs.data$ps1)
  ATT.ipw.c.bs[b]<-mean(bs.data$y[bs.data$treat==1])- mean(bs.data$y*temp.oddc1)
  ATT.ipw.n.c.bs[b]<-mean(bs.data$y[bs.data$treat==1])- sum(bs.data$y*temp.oddc1)/sum(temp.oddc1)

#################################################################
########## Weighting methods for ATE with incorrect PS
#################################################################
   treatnum1<- sum(bs.data$y*bs.data$treat/(bs.data$psic))
   treatdenom1<- sum(bs.data$treat/bs.data$psic)
   contnum1<- sum(bs.data$y*(1-bs.data$treat)/(1-bs.data$psic))
   contdenom1<- sum((1-bs.data$treat)/(1-bs.data$psic))
   ATE.ipw.n.ic.bs[b]<- (treatnum1/treatdenom1)-(contnum1/contdenom1)
   ATE.ipw.ic.bs[b]<- treatnum1/N-contnum1/N

#################################################################
########## Weighting methods for ATE with correct PS
#################################################################

   treatnum2<- sum(bs.data$y*bs.data$treat/(bs.data$ps1))
   treatdenom2<- sum(bs.data$treat/bs.data$ps1)
```

```
    contnum2<- sum(bs.data$y*(1-bs.data$treat)/(1-bs.data$ps1))
    contdenom2<- sum((1-bs.data$treat)/(1-bs.data$ps1))
    ATE.ipw.n.c.bs[b]<- (treatnum2/treatdenom2)-(contnum2/contdenom2)
    ATE.ipw.c.bs[b]<- treatnum2/N-contnum2/N


  }
  SE.ATT.ipw.ic<-sd(ATT.ipw.ic.bs)
  SE.ATT.ipw.ic.n  <-sd(ATT.ipw.n.ic.bs)
  SE.ATT.ipw.c<-sd( ATT.ipw.c.bs)
  SE.ATT.ipw.n.c<-sd( ATT.ipw.n.c.bs)
  SE.ATE.ipw.n.ic<-sd(ATE.ipw.n.ic.bs)
  SE.ATE.ipw.ic<-sd(ATE.ipw.ic.bs)
  SE.ATE.ipw.n.c<-sd(ATE.ipw.n.c.bs)
  SE.ATE.ipw.c<-sd(ATE.ipw.c.bs)



  return(list( SE.ATT.ipw.ic,SE.ATT.ipw.ic.n,SE.ATT.ipw.c,SE.ATT.ipw.n.c,
                          SE.ATE.ipw.n.ic,SE.ATE.ipw.ic,SE.ATE.ipw.n.c,SE.ATE.ipw.c))
}

#############################################################################
#############################################################################
## Fixing beta for the T.OR
#############################################################################
#############################################################################
beta0<- 0
beta1<- 2
beta2<- 3
beta3<- 2
beta4<- -4


TATE<- 2 ## True ATE
SNR<- 50 ## Signal noise ratio
set.seed(9999)## Setting the seed
n<- 10 ## Total number of simulations

#############################################################################
#############################################################################
## Defining the variables to store the values of ATT, ATE, SE for ATT,
## SE for ATE by using the different methods Weighted. ASMD for ATT using F.PS,
##Weighted ASMD for ATE using F.PS, Weighted ASMD for ATE using T.PS,
## Weighted ASMD for ATT using T.PS, ASMD for stratification using F.PS,
## ASMD for stratification using T.PS
#############################################################################
#############################################################################


TATT<-ATT.cov<-ATT.reg<-ATT.regc<-ATT.strat<-ATT.stratc<-rep(NA,n)
ATT.ipw.ic<-ATT.ipw.n.ic<-ATT.ipw.c<-ATT.ipw.n.c<-rep(NA,n)
SE.ATT.cov<-SE.ATT.reg<-SE.ATT.regc<-SE.ATT.strat<-SE.ATT.stratc<-rep(NA,n)

TATE<-ATE.cov<-ATE.reg<-ATE.regc<-ATE.strat<-ATE.stratc<-rep(NA,n)
ATE.ipw.n.ic<-ATE.ipw.ic<-ATE.ipw.n.c<-ATE.ipw.c<- rep(NA,n)
ATEDRCOR<-ATEDRCPS<-ATEDRIC<-rep(NA,n)
SE.ATE.cov<- SE.ATE.reg<-SE.ATE.regc<-rep(NA,n)
SE.ATE.strat<-SE.ATE.stratc<-rep(NA,n)

SE.ATT.ipw.ic<-SE.ATT.ipw.ic.n<-SE.ATT.ipw.c<-rep(NA,n)
SE.ATT.ipw.n.c<-SE.ATE.ipw.n.ic<-SE.ATE.ipw.ic<-rep(NA,n)
SE.ATE.ipw.n.c<-SE.ATE.ipw.c<-rep(NA,n)
SEDRCOR<-SEDRCPS<-SEDRIC<-rep(NA,n)

ASMDATEB<-ASMDATE<- ASMDATT<-rep(NA,n)
ASMDATEC<-ASMDATTC<-ASMDATTS<- rep(NA,n)
ASMDATES<- ASMDATTSC<-ASMDATESC<-matrix(NA,n,10)

#############################################################################

##sim=1
for(sim in 1:n){
N<- 1000  ## Sample Size

#############################################################################
#############################################################################
## Generating the dataset for normal distribution
## with mean zero and variance 1
#############################################################################
#############################################################################
x1<- rnorm(N,0,1)
x2<- rnorm(N,0,1)
x3<- rnorm(N,0,1)
x4<- rnorm(N,0,1)
x5<- rnorm(N,0,1)
x6<- rnorm(N,0,1)
x7<- rnorm(N,0,1)
x8<- rnorm(N,0,1)
x9<- rnorm(N,0,1)
x10<- rnorm(N,0,1)
```

```
x1.sq<-(x1)^2
x2.sq<-(x2)^2
x.23<- x2*x3
x.345<- x3*x4*x5

#####################################################################
#####################################################################
## betas for treatment selection model
#####################################################################
#####################################################################

beta.low <- log(1.25)
beta.med <- log(1.5)
beta.high <- log(1.75)
beta.v.high <- log(2)
beta.0.treat<- -1.396

#####################################################################
#####################################################################
## Generating the treatment variable
#####################################################################
#####################################################################

logit.treat<- beta.0.treat+beta.low*x1 + beta.med*x2 + beta.high*x3
                          + beta.low*x1.sq + beta.med*x2.sq
                                  + beta.high*x.23+ beta.v.high*x.345
p.treat <- exp(logit.treat)/(1 + exp(logit.treat))
treat <- rbinom(N,1,p.treat)
sum(treat)

#####################################################################
## Estimating the outcome variable
#####################################################################
#####################################################################
sigma<- sqrt(var(ey)/SNR)
y<- ey + rnorm(N, 0, sigma)


## Now creating a dataset
datasim<- data.frame(y,treat,x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x1.sq, x2.sq, x.23, x.345)
#####################################################################
#####################################################################
## Creating the potential outcomes for each subject
#####################################################################
#####################################################################
datasim$y1<-beta0+ beta1*x1.sq + beta2*x4
datasim$y0<-  beta0 + beta3*x1*x4+ beta4*x5
##TATE[sim]<- mean(datasim$y1 - datasim$y0)
##TATT[sim]<- mean(datasim$y1[treat==1] - datasim$y0[treat==1])

#####################################################################
#####################################################################
## Estimating the propensity scores by using the F.PS
#####################################################################
#####################################################################

modelps=glm(treat~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10,family=binomial, data = datasim);
datasim$ps=predict(modelps,type="response")

#####################################################################
#####################################################################
## Estimating the propensity scores by using the T.PS
#####################################################################
#####################################################################

modelpsc<- glm(treat~x1+x2+x3+x1.sq+x2.sq+x.23+x.345,family=binomial, data=datasim)
datasim$psc=predict(modelpsc,type="response")
#####################################################################
#####################################################################
## Estimating the ASMD using the propensity scores estimated by F.PS
#####################################################################
#####################################################################

## Estimating the covariate balance using the ASMD using incorrect PS.
## Covariate balance when estimating ATE
dataX<- subset(datasim,select=c(3:12))
datagroup1<- subset(datasim, datasim$treat== 1,select =c(3:12,19))
datagroup0<- subset(datasim, datasim$treat== 0,select =c(3:12,19))
xbar1ATE<- apply(datagroup1,2,function(a,b)weighted.mean(a,b),b=1/datagroup1$ps);
                xbar1ATE<- t(as.matrix(unname(xbar1ATE[-11])));
xbar0ATE<- apply(datagroup0,2,function(a,b)weighted.mean(a,b),b=1/(1-datagroup0$ps));
                xbar0ATE<- t(as.matrix(unname(xbar0ATE[-11])));
xbar1ATEB<- apply(datagroup1,2,mean); xbar1ATEB<- t(as.matrix(unname(xbar1ATEB[-11])));
xbar0ATEB<- apply(datagroup0,2,mean); xbar0ATEB<- t(as.matrix(unname(xbar0ATEB[-11])));

sdATE<-t(as.matrix(unname(apply(dataX, 2, sd))))
 ## Calculate the ASMD using the F.PS when ATE is to be estimated
ASMDATE[sim,]<- abs(xbar1ATE-xbar0ATE)/sdATE
## Calculate the ASMD before balancing the covariates,
ASMDATEB[sim,]<- abs(xbar1ATEB-xbar0ATEB)/sdATE
```

```
## Covariate balance when estimating ATT
xbar1ATT<- apply(datagroup1,2,function(a,b)weighted.mean(a,b),b=rep(1,length(datagroup1$x1)));
                xbar1ATT<- t(as.matrix(unname(xbar1ATT[-11])));
xbar0ATT<- apply(datagroup0,2,function(a,b)weighted.mean(a,b),b=(datagroup0$ps)/(1-datagroup0$ps));
                xbar0ATT<- t(as.matrix(unname(xbar0ATT[-11])));
sdATT<-apply(datagroup1, 2, sd); sdATT<-sdATT[-11];
## Calculate the ASMD using the F.PS when ATT is to be estimated
ASMDATT[sim,]<- abs(xbar1ATT-xbar0ATT)/sdATT


#################################################################
#################################################################
## Estimating the ASMD using the propensity scores estimated by T.PS
#################################################################
#################################################################

##Estimating the ASMD using the correct PS
datagroup1C<- subset(datasim, datasim$treat== 1,select =c(3:12,20))
datagroup0C<- subset(datasim, datasim$treat== 0,select =c(3:12,20))

xbar1ATEC<- apply(datagroup1C,2,function(a,b)weighted.mean(a,b),b=1/datagroup1C$psc);
        xbar1ATEC<- t(as.matrix(unname(xbar1ATEC[-11])));
xbar0ATEC<- apply(datagroup0C,2,function(a,b)weighted.mean(a,b),b=1/(1-datagroup0C$psc));
                xbar0ATEC<- t(as.matrix(unname(xbar0ATEC[-11])));
sdATE<-t(as.matrix(unname(apply(dataX, 2, sd))))
ASMDATEC[sim,]<- abs(xbar1ATEC-xbar0ATEC)/sdATE

xbar1ATTC<- apply(datagroup1C,2,function(a,b)weighted.mean(a,b),b=rep(1,length(datagroup1C$x1)));
         xbar1ATTC<- t(as.matrix(unname(xbar1ATTC[-11])));
xbar0ATTC<- apply(datagroup0C,2,function(a,b)weighted.mean(a,b),b=(datagroup0C$psc)/(1-datagroup0C$psc));
        xbar0ATTC<- t(as.matrix(unname(xbar0ATTC[-11])));
sdATT<-apply(datagroup1, 2, sd); sdATT<-sdATT[-11];
## Calculate the ASMD using the F.PS when ATT is to be estimated
ASMDATTC[sim,]<- abs(xbar1ATTC-xbar0ATTC)/sdATT

#################################################################
#################################################################
## Estimating the ASMD ATT by different methods
#################################################################
#################################################################


#########################################################
##    Regression for ATT using False PS
#########################################################

modelglm<- lm(y~ps*treat, data= datasim)

#########################################################
### ATT when the mean of the PS for the treated group is used

## Provides the ATT estimate via regression using false PS.
ATT.reg[sim]<-modelglm$coef[3]+modelglm$coef[4]*mean(datasim$ps[datasim$treat==1])

#################################################################
### Estimating the standard error for ATT when mean of PS for treated group
### is used
#################################################################

Xtreatmean1<- matrix(c(1, mean(datasim$ps[datasim$treat==1])), nrow=1, ncol=2)
### Provides the SE for ATT estimated via regression
SE.ATT.reg[sim]<- sqrt(Xtreatmean1%*%cov%*%t(Xtreatmean1))

#########################################################
## Regression for ATT using the Correct PS
#########################################################
modelglmc<- lm(y~psc*treat, data= datasim)
## Provides the ATT estimate via regression using false PS.

ATT.regc[sim]<-modelglmc$coef[3]+modelglmc$coef[4]*mean(datasim$psc[datasim$treat==1])
#################################################################
## Estimating the Variance for ATT When estimated using correct PS as covariate
#################################################################
varcovc<- vcov(modelglmc)
covc<- matrix(c(varcovc[3,3], varcovc[3,4], varcovc[4,3], varcovc[4,4]),nrow=2, ncol=2)

Xtreatmean1c<- matrix(c(1, mean(datasim$psc[datasim$treat==1])), nrow=1, ncol=2)
### Provides the SE for ATT estimated via regression
SE.ATT.regc[sim]<- sqrt(Xtreatmean1c%*%covc%*%t(Xtreatmean1c))


#########################################################
### Estimating the ATT using startification using False PS
#########################################################
breakvals<- as.numeric(quantile(datasim$ps, c(0.40,0.60,0.80)))
datasim$strata1<-cut(datasim$ps,br=c(0,breakvals,1), right=FALSE, labels=c(1:4))
a<- table(datasim$strata1)
a1<- table(datasim$strata1[treat==1])
fits <- lmList( y ~ treat | strata1, data=datasim)
## Gives the ATT estimating using the stratification method when using F.PS
```

```
ATT.strat[sim]<-sum(a1/sum(a1)*summary(fits)$coefficients[,1,2])
#################################################################
### Estimating the SE for ATT stratification
#################################################################
## Gives the ATT estimating using the stratification method when using F.PS
SE.ATT.strat[sim]<-sqrt(sum((a1/sum(a1))^2*(summary(fits)$coefficients[,2,2])^2))

#################################################################
### Estimating the ATT using startification using Correct PS
#################################################################
breakvalsc<- as.numeric(quantile(datasim$psc, c(0.40,0.60,0.80)))
datasim$strata1c<-cut(datasim$psc,br=c(0,breakvalsc,1), right=FALSE, labels=c(1:4))
ac<- table(datasim$strata1c)
a1c<- table(datasim$strata1c[treat==1])
fitsc <- lmList( y ~ treat | strata1c, data=datasim)
## Gives the ATT estimating using the stratification method when using F.PS
ATT.stratc[sim]<-sum(a1c/sum(a1c)*summary(fitsc)$coefficients[,1,2])

#################################################################
### Estimating the Standard error for stratification method using the True PS
#################################################################
SE.ATT.stratc[sim]<-sqrt(sum((a1c/sum(a1c))^2*(summary(fitsc)$coefficients[,2,2])^2))

#################################################################
#################################################################
## Estimating the ASMD for stratification when the PS is estimated using the F.PS
#################################################################
#################################################################

datat1<- subset(datasim, select=c(3:12,21), datasim$treat== 1)
datat0<- subset(datasim, select=c(3:12,21), datasim$treat== 0)
meant1s<-aggregate(datat1[1:10], by=list(datat1$strata1),FUN=mean)[,-1]
meant0s<-aggregate(datat0[1:10], by=list(datat0$strata1),FUN=mean)[,-1]
meant1sn<-as.matrix(meant1s)
meant0sn<-as.matrix(meant0s)
dms<-abs(meant1sn-meant0sn)
dmsm<-apply(dms,2,mean)
sdtrt<- apply(datat1,2,sd)[-11]

ASMDATTS[sim,]<-dmsm/sdtrt ## ASMD after stratification for ATT
ASMDATES[sim,]<-dmsm/sdATE  ## ASMD after stratification for ATE

#################################################################
#################################################################
## Estimating the ASMD for stratification when the PS is estimated using the T.PS
#################################################################
#################################################################

datat1c<- subset(datasim, select=c(3:12,22), datasim$treat== 1)
datat0c<- subset(datasim, select=c(3:12,22), datasim$treat== 0)
meant1sc<-aggregate(datat1c[1:10], by=list(datat1c$strata1c),FUN=mean)[,-1]
meant0sc<-aggregate(datat0c[1:10], by=list(datat0c$strata1c),FUN=mean)[,-1]
meant1snc<-as.matrix(meant1sc)
meant0snc<-as.matrix(meant0sc)
dmsc<-abs(meant1snc-meant0snc)
dmsmc<-apply(dmsc,2,mean)
ASMDATTSC[sim,]<-dmsmc/sdtrt ## ASMD after stratification for ATT using T.PS
ASMDATESC[sim,]<-dmsmc/sdATE ## ASMD after stratification for ATE using T.PS



#################################################################
#################################################################
##  Weight for ATT using F.PS
#################################################################
temp.odd<-(1-datasim$treat)*datasim$ps/(1-datasim$ps)
## ATT estimator for IPW
ATT.ipw.ic[sim]<-mean(datasim$y[datasim$treat==1])- mean(datasim$y*temp.odd)
## ATT estimator for IPWN
ATT.ipw.n.ic[sim]<-mean(datasim$y[datasim$treat==1])- sum(datasim$y*temp.odd)/sum(temp.odd)

#################################################################
#########################################################################
##  Weight for ATT using correct model T.PS
#################################################################
temp.oddc<-(1-datasim$treat)*datasim$psc/(1-datasim$psc)
## ATT estimator for IPW
ATT.ipw.c[sim]<-mean(datasim$y[datasim$treat==1])- mean(datasim$y*temp.oddc)
## ATT estimator for IPWN
ATT.ipw.n.c[sim]<-mean(datasim$y[datasim$treat==1])- sum(datasim$y*temp.oddc)/sum(temp.oddc)

#################################################################
#################################################################
### ATE Estimates using different methods
#################################################################
#################################################################

#########################################################
## Estimating ATE using the mean of ps for whole sample using False PS
#########################################################
## Gives the ATE estimator via regression using F.PS
```

```
ATE.reg[sim]<-modelglm$coef[3]+modelglm$coef[4]*mean(datasim$ps)

#########################################################
### Estimating the se for ATE when the mean of ps is used (False PS)
Xpopmean1<- matrix(c(1, mean(datasim$ps)), nrow=1, ncol=2)
## Gives the SE of ATE estimator via regression using F.PS
SE.ATE.reg[sim]<- sqrt(Xpopmean1%*%cov%*%t(Xpopmean1))

#########################################################
###### Estimating ATE using the mean of ps for whole sample using True PS
#########################################################
## Gives the ATE estimator via regression using T.PS

ATE.regc[sim]<-modelglmc$coef[3]+modelglmc$coef[4]*mean(datasim$psc)
Xpopmean1c<- matrix(c(1, mean(datasim$psc)), nrow=1, ncol=2)
## Gives the SE of ATE estimator via regression using F.PS
SE.ATE.regc[sim]<- sqrt(Xpopmean1c%*%covc%*%t(Xpopmean1c))



#################################################
### Finding ATE using the stratification using False PS
#########################################################
## Gives the ATE estimator via stratification using F.PS
ATE.strat[sim]<-sum(a/N*summary(fits)$coefficients[,1,2])

#################################################
## Finding the SE for ATE stratification
#########################################################
## Gives the SE of ATE estimator via stratification using F.PS

SE.ATE.strat[sim]<-sqrt(sum((a/N)^2*(summary(fits)$coefficients[,2,2])^2))

#########################################################
## Finding ATE using the stratification using True PS
#########################################################
## Gives the ATE estimator via stratification using TPS
ATE.stratc[sim]<-sum(ac/N*summary(fitsc)$coefficients[,1,2])
SE.ATE.stratc[sim]<-sqrt(sum((ac/N)^2*(summary(fitsc)$coefficients[,2,2])^2))

#########################################################
#########################################################
##  IPW for ATE using incorrect PS
#########################################################
treatnum<- datasim$treat*datasim$y/datasim$ps
treatdenom<- datasim$treat/datasim$ps
contnum<- (1-datasim$treat)*datasim$y/(1-datasim$ps)
contdenom<- (1-datasim$treat)/(1-datasim$ps)

## Provides the ATE value via IPWN using F.PS
ATE.ipw.n.ic[sim]<- sum(treatnum)/sum(treatdenom) - sum(contnum)/sum(contdenom)
  ## Provides the ATE value via IPW using F.PS
ATE.ipw.ic[sim]<- mean(treatnum)-mean(contnum)

#########################################################
#########################################################
##  IPW for ATE using correct PS
#########################################################
treatnumc<- datasim$treat*datasim$y/datasim$psc
treatdenomc<- datasim$treat/datasim$psc
contnumc<- (1-datasim$treat)*datasim$y/(1-datasim$psc)
contdenomc<- (1-datasim$treat)/(1-datasim$psc)
## Provides the ATE value via IPWN using T.PS
ATE.ipw.n.c[sim]<- sum(treatnumc)/sum(treatdenomc) - sum(contnumc)/sum(contdenomc)
 ## Provides the ATE value via IPW using F.PS

ATE.ipw.c[sim]<- mean(treatnumc)-mean(contnumc)



#########################################################
### ATE using DR incorrect PS model, but correct OR
#########################################################
datasim$x1.sq.t<- x1.sq*treat
datasim$x4.t<- x4*treat
datasim$x1x4t<- x1*x4*(1-treat)
datasim$x1x4<- x1*x4
datasim$x5t<- x5*(1-treat)

## Finding the correct multiple regression and finding the predicted values
modelc<- lm(y~x1.sq.t+x4.t+x1x4t+x5t, data = datasim)
## Now finding the predicting values

datasim$treat1c<- modelc$coeff[1]+ as.matrix(datasim[c("x1.sq", "x4")])%*%as.vector(modelc$coeff[c(2,3)])
datasim$cont1c<- modelc$coeff[1]+ as.matrix(datasim[c("x1x4", "x5")])%*%as.vector(modelc$coeff[c(4,5)])
## Provides the ATE estimator via DR with TOR
ATEDRCOR[sim]<- mean((datasim$treat*datasim$y - (datasim$treat-datasim$ps)*datasim$treat1c)/datasim$ps)-
            mean(((1-datasim$treat)*datasim$y+(datasim$treat-datasim$ps)*datasim$cont1c)/(1-datasim$ps))

## Now estimating the SE for DR when incorrect PS model is used
```

```
datasim$IDRCOR<-((datasim$treat*datasim$y) - datasim$treat1c*(datasim$treat-datasim$ps))/(datasim$ps)
                        - (((1-datasim$treat)*datasim$y
                        +datasim$cont1c*(datasim$treat-datasim$ps))/(1-datasim$ps))- ATEDRCOR[sim]
SEDRCOR[sim]<-sqrt(1/(N^2)*sum((datasim$IDRCOR)^2))

#############################################################################
## ATE using DR incorrect OR model, but correct PS model
#############################################################################

modelic<- lm(y~treat+x1+x2+x3+x4+x5+x6+x7+x8+x9+x10, data=datasim)
datasim$treat1ic<-modelic$coeff[1]+modelic$coeff[2]+
 as.matrix(datasim[c("x1","x2","x3", "x4","x5","x6","x7","x8","x9","x10")])%*%as.vector(modelic$coeff[c(-1,-2)])
datasim$cont1ic<- modelic$coeff[1]
+ as.matrix(datasim[c("x1","x2","x3", "x4","x5","x6","x7","x8","x9","x10")])%*%as.vector(modelic$coeff[c(-1,-2)])

ATEDRCPS[sim]<- mean((datasim$treat*datasim$y - (datasim$treat-datasim$psc)*datasim$treat1ic)/datasim$psc)-
     mean(((1-datasim$treat)*datasim$y+(datasim$treat-datasim$psc)*datasim$cont1ic)/(1-datasim$psc))

## Now estimating the SE for DR when correct PS model is used
datasim$IDRCPS<-((datasim$treat*datasim$y) - datasim$treat1ic*(datasim$treat-datasim$psc))/(datasim$psc)
                        - (((1-datasim$treat)*datasim$y
                        +datasim$cont1ic*(datasim$treat-datasim$psc))/(1-datasim$psc))- ATEDRCPS[sim]
SEDRCPS[sim]<-sqrt(1/(N^2)*sum((datasim$IDRCPS)^2))

#############################################################################
## ATE  using DR incorrect PS and incorrect OR model
#############################################################################

ATEDRIC[sim]<- mean((datasim$treat*datasim$y - (datasim$treat-datasim$ps)*datasim$treat1ic)/datasim$ps)-
     mean(((1-datasim$treat)*datasim$y+(datasim$treat-datasim$ps)*datasim$cont1ic)/(1-datasim$ps))

## Now estimating the SE for DR when both OR and PS models are incorrect

datasim$IDRIC<-((datasim$treat*datasim$y) - datasim$treat1ic*(datasim$treat-datasim$ps))/(datasim$ps)
                        - (((1-datasim$treat)*datasim$y+datasim$cont1ic*(datasim$treat-datasim$ps))/(1-datasim$ps))
SEDRIC[sim]<-sqrt(1/(N^2)*sum((datasim$IDRIC)^2))


datasim1<- datasim[c(1:16)]
## Function to estimate the SE when ATT and ATE are estimated using IPW and IPWN
SE.result<-Var.bootstrap(dataset=datasim1,  nB=200)


SE.ATT.ipw.ic[sim]<- SE.result[[1]]
SE.ATT.ipw.ic.n[sim]<-SE.result[[2]]
SE.ATT.ipw.c[sim]<-SE.result[[3]]
SE.ATT.ipw.n.c[sim]<-SE.result[[4]]
SE.ATE.ipw.n.ic[sim]<-SE.result[[5]]
SE.ATE.ipw.ic[sim]<-SE.result[[6]]
SE.ATE.ipw.n.c[sim]<-SE.result[[7]]
SE.ATE.ipw.c[sim]<-SE.result[[8]]


print(sim)
}

#############################################################################
#############################################################################
## Estimating the mean of ATT and ATE for each method, Bias, average
## of standard error, the ESE and the RMSE for table 2.3
## When we have logistic regression
#############################################################################
#############################################################################

ATATT<- 2.485
ATATE<- 2
#####################################################
### Estimates of the ATT
#####################################################
##ATT.COV<- mean(ATT.cov)
ATT.REG<- mean(ATT.reg);                    BIAS.ATT.REG<-ATT.REG - ATATT;
ATT.REG.C<-mean(ATT.regc);                  BIAS.ATT.REG.C<-ATT.REG.C - ATATT;
ATT.STRAT<- mean(ATT.strat);                BIAS.ATT.STRAT<-ATT.STRAT - ATATT;
ATT.STRAT.C<-mean(ATT.stratc);   BIAS.ATT.STRAT.C<-ATT.STRAT.C - ATATT;
ATT.IPW.IC<- mean(ATT.ipw.ic);   BIAS.ATT.IPW.IC<-ATT.IPW.IC - ATATT;
ATT.IPW.N.IC<- mean(ATT.ipw.n.ic);          BIAS.ATT.IPW.N.IC<-ATT.IPW.N.IC - ATATT;
ATT.IPW.C<- mean(ATT.ipw.c);                BIAS.ATT.IPW.C<-ATT.IPW.C - ATATT;
ATT.IPW.N.C<- mean(ATT.ipw.n.c);            BIAS.ATT.IPW.N.C<-ATT.IPW.N.C - ATATT;

#####################################################
## SE of the ATT
#####################################################

##SE.ATT.COV<- mean(SE.ATT.cov)
SE.ATT.REG<- mean(SE.ATT.reg)
SE.ATT.REG.C<-mean(SE.ATT.regc)
SE.ATT.STRAT<- mean(SE.ATT.strat)
SE.ATT.STRAT.C<- mean(SE.ATT.stratc)
SE.ATT.IPW.IC<- mean(SE.ATT.ipw.ic)
SE.ATT.IPW.N.IC<- mean(SE.ATT.ipw.ic.n)
```

```
SE.ATT.IPW.C<- mean(SE.ATT.ipw.c)
SE.ATT.IPW.N.C<- mean(SE.ATT.ipw.n.c)

#######################################################
### Empirical SE for the estimates
#######################################################

ESE.ATT.REG<- sd(ATT.reg)
ESE.ATT.REG.C<-sd(ATT.regc)
ESE.ATT.STRAT<- sd(ATT.strat)
ESE.ATT.STRAT.C<-sd(ATT.stratc)
ESE.ATT.IPW.IC<- sd(ATT.ipw.ic)
ESE.ATT.IPW.N.IC<- sd(ATT.ipw.n.ic)
ESE.ATT.IPW.C<- sd(ATT.ipw.c)
ESE.ATT.IPW.N.C<- sd(ATT.ipw.n.c)

RMSE.ATT.REG<- sqrt(mean((ATT.reg-ATATT)^2))
RMSE.ATT.REGC<- sqrt(mean((ATT.regc-ATATT)^2))

RMSE.ATT.STRAT<- sqrt(mean((ATT.strat-ATATT)^2))
RMSE.ATT.STRATC<- sqrt(mean((ATT.stratc-ATATT)^2))
RMSE.ATT.IPW.IC<- sqrt(mean((ATT.ipw.ic-ATATT)^2))
RMSE.ATT.IPWN.IC<- sqrt(mean((ATT.ipw.n.ic -ATATT)^2))
RMSE.ATT.IPW.C<- sqrt(mean((ATT.ipw.c - ATATT)^2))
RMSE.ATT.IPWN.C<- sqrt(mean((ATT.ipw.n.c -ATATT)^2))

#######################################################
## Estimate of the ATE
#######################################################
##ATE.COV<- mean(ATE.cov);

ATE.REG<- mean(ATE.reg);                    BIAS.ATE.REG<-ATE.REG - ATATE;
ATE.REG.C<-mean(ATE.regc);                  BIAS.ATE.REG.C<-ATE.REG.C - ATATE;
ATE.STRAT<- mean(ATE.strat);                BIAS.ATE.STRAT<-ATE.STRAT - ATATE;
ATE.STRAT.C<-mean(ATE.stratc);   BIAS.ATE.STRAT.C<-ATE.STRAT.C - ATATE;
ATE.IPW.IC<- mean(ATE.ipw.ic);   BIAS.ATE.IPW.IC<-ATE.IPW.IC - ATATE;
ATE.IPW.N.IC<- mean(ATE.ipw.n.ic);          BIAS.ATE.IPW.N.IC<-ATE.IPW.N.IC - ATATE;
ATE.IPW.C<- mean(ATE.ipw.c);                BIAS.ATE.IPW.C<-ATE.IPW.C - ATATE;
ATE.IPW.N.C<- mean(ATE.ipw.n.c);            BIAS.ATE.IPW.N.C<-ATE.IPW.N.C - ATATE;
ATE.DR.COR<-mean(ATEDRCOR);                 BIAS.ATE.DR.COR<-ATE.DR.COR - ATATE;
ATE.DR.CPS<-mean(ATEDRCPS);                 BIAS.ATE.DR.CPS<-ATE.DR.CPS - ATATE;
ATE.DR.IC <- mean(ATEDRIC);                 BIAS.ATE.DR.IC<-ATE.DR.IC - ATATE;

#######################################################
## SE of the ATE
#######################################################

##SE.ATE.COV<- mean(SE.ATE.cov)
SE.ATE.REG<- mean(SE.ATE.reg)
SE.ATE.REG.C<-mean(SE.ATE.regc)
SE.ATE.STRAT<- mean(SE.ATE.strat)
SE.ATE.STRAT.C<-mean(SE.ATE.stratc)
SE.ATE.IPW.IC<- mean(SE.ATE.ipw.ic)
SE.ATE.IPW.N.IC<- mean(SE.ATE.ipw.n.ic)
SE.ATE.IPW.C<- mean(SE.ATE.ipw.c)
SE.ATE.IPW.N.C<- mean(SE.ATT.ipw.n.c)
SE.ATE.DR.COR<- mean (SEDRCOR)
SE.ATE.DR.CPS<- mean (SEDRCPS)
SE.ATE.DR.IC<- mean (SEDRIC)

#######################################################
## ESE of the ATE
#######################################################
ESE.ATE.REG.C<- sd(ATE.regc)
ESE.ATE.REG<- sd(ATE.reg)
ESE.ATE.STRAT.C<- sd(ATE.stratc)
ESE.ATE.STRAT<- sd(ATE.strat)
ESE.ATE.IPW.IC<- sd(ATE.ipw.ic)
ESE.ATE.IPW.N.IC<- sd(ATE.ipw.n.ic)
ESE.ATE.IPW.C<- sd(ATE.ipw.c)
ESE.ATE.IPW.N.C<- sd(ATE.ipw.n.c)
ESE.ATE.DR.COR<- sd(ATEDRCOR)
ESE.ATE.DR.CPS<- sd(ATEDRCPS)
ESE.ATE.DR.IC<- sd(ATEDRIC)

RMSE.ATE.REG<- sqrt(mean((ATE.reg-ATATE)^2))
RMSE.ATE.REG.C<- sqrt(mean((ATE.regc-ATATE)^2))

RMSE.ATE.STRAT<- sqrt(mean((ATE.strat-ATATE)^2))
RMSE.ATE.STRAT.C<- sqrt(mean((ATE.stratc-ATATE)^2))
RMSE.ATE.IPW.IC<- sqrt(mean((ATE.ipw.ic-ATATE)^2))
RMSE.ATE.IPWN.IC<-sqrt(mean((ATE.ipw.n.ic-ATATE)^2))
RMSE.ATE.IPW.C<- sqrt(mean((ATE.ipw.c-ATATE)^2))
RMSE.ATE.IPWN.C<-sqrt(mean((ATE.ipw.n.c-ATATE)^2))
RMSE.ATE.DR.COR<-sqrt(mean((ATEDRCOR-ATATE)^2))
RMSE.ATE.DR.CPS<-sqrt(mean((ATEDRCPS-ATATE)^2))
RMSE.ATE.DR.IC<-sqrt(mean((ATEDRIC-ATATE)^2))
```

```
#############################################################################
#############################################################################
## R Code producing results for chapter 2 when the propensity score
##is estimated using the logistic regression and variables are
## dependently normally distributed.
#############################################################################
#############################################################################

## The R Code for dependently normally distributed is same as of independent variables.
## We just need to add an extra package library(mvtnorm)
and the data is to be generated as following

## Creating Matrix for mean and variance covariance Matrix
meanmat<- rep(0,10)
m<- diag(1,10,10)
m[lower.tri(m)] <- 0.50
m[upper.tri(m)] <- 0.50

x <- rmvnorm(n=N, mean=meanmat, sigma=m)
x1<- x[,1]
x2<- x[,2]
x3<- x[,3]
x4<- x[,4]
x5<- x[,5]
x6<- x[,6]
x7<- x[,7]
x8<- x[,8]
x9<- x[,9]
x10<- x[,10]
x1.sq<- (x1)^2
x2.sq<- (x2)^2
x.23<- x2*x3
x.345<-x3*x4*x5

## The results for ATT and ATE can be estimated by using the code given above.
```

```
#############################################################################
#############################################################################
## R Code producing results for chapter 2 when the propensity score
##is estimated using the lGBM and variables are
## dependently normally distributed.
#############################################################################
#############################################################################

#############################################################################
#############################################################################
## Commands needed to run R Code on the Parallel computing environment
#############################################################################
#############################################################################

##args <- commandArgs()
## start w/args[3]
##Run <- as.numeric(args[3])
##Sim <- as.numeric(args[4])
##name <- paste("RESULT", Run,Sim, "RData", sep = ".")

#############################################################################
#############################################################################
## Packages needed to run the code
#############################################################################
#############################################################################
library(gbm)
library(survival)
library(splines)
library(lattice)
library(parallel)
library(lme4)
library(Matrix)
library(Rcpp)

#############################################################################
#############################################################################
## Function to estimate standard error via bootstrap for IPW Methods
#############################################################################
#############################################################################
Var.bootstrap<- function(dataset, depth,  nB=200)
{
  ATE.cov.bs<- ATT.cov.bs<- ATE.strata.bs<- ATT.strata.bs<-rep(NA,nB)
ATE.ipw.bs<-ATE.ipw.n.bs<-ATEDRC.bs<-ATEDRIC.bs<-rep(NA,nB)
```

```
  ATT.ipw.bs<-ATT.ipw.n.bs<-ATT.reg.bs<-ATE.reg.bs<-rep(NA,nB)
  N<-nrow(dataset)
  # b<-1
  for(b in 1:nB){
    bs.data<-as.data.frame(dataset[sample(nrow(dataset),N, replace=TRUE),])
    ps.logit1 <- gbm(treat~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10, data = bs.data,
                     distribution = "bernoulli",
                     var.monotone = NULL,
                     n.trees = 9000,
                     interaction.depth = depth,
                     n.minobsinnode = 10,
                     shrinkage = 0.05,
                     bag.fraction = 1.0,
                     train.fraction = 1.0,
                     cv.folds=10,
                     keep.data = TRUE,
                     verbose = FALSE,
                     n.cores = NULL)

    bs.data$psbs<-predict(ps.logit1, data=bs.data,type="response")



    ######################################################################
    ########## Weighting methods for ATE
    ######################################################################

    treatnum1<- sum(bs.data$y*bs.data$treat/(bs.data$psbs))
    treatdenom1<- sum(bs.data$treat/bs.data$psbs)
    contnum1<- sum(bs.data$y*(1-bs.data$treat)/(1-bs.data$psbs))
    contdenom1<- sum((1-bs.data$treat)/(1-bs.data$psbs))
    ATE.ipw.n.bs[b]<- (treatnum1/treatdenom1)-(contnum1/contdenom1)
    ATE.ipw.bs[b]<- treatnum1/N-contnum1/N

    ######################################################################
    ## Now estimating the ATE using the DR with correct OR
    ######################################################################
    bs.data$x1.sq.t<- bs.data$x1.sq*bs.data$treat
    bs.data$x4.t<- bs.data$x4*bs.data$treat
    bs.data$x1x4t<- bs.data$x1*bs.data$x4*(1-bs.data$treat)
    bs.data$x1x4<- bs.data$x1*bs.data$x4
    bs.data$x5t<- bs.data$x5*(1-bs.data$treat)

    ## Finding the correct multiple regression and finding the predicted values
    modelc1<- lm(y~x1.sq.t+x4.t+x1x4t+x5t, data = bs.data)
    ## Now finding the predicting values

    bs.data$treat1c1<- modelc1$coeff[1]+ as.matrix(bs.data[c("x1.sq", "x4")])%*%as.vector(modelc1$coeff[c(2,3)])
    bs.data$cont1c1<- modelc1$coeff[1]+ as.matrix(bs.data[c("x1x4", "x5")])%*%as.vector(modelc1$coeff[c(4,5)])

    ATEDRC.bs[b]<- mean((bs.data$treat*bs.data$y - (bs.data$treat-bs.data$psbs)*bs.data$treat1c1)/bs.data$psbs)-
      mean(((1-bs.data$treat)*bs.data$y+(bs.data$treat-bs.data$psbs)*bs.data$cont1c1)/(1-bs.data$psbs))


    ######################################################################
    ##### Now estimating the ATE using the DR with Incorrect OR
    ######################################################################
    modelic1<- lm(y~treat+x1+x2+x3+x4+x5+x6+x7+x8+x9+x10, data=bs.data)
    bs.data$treat1ic1<-modelic1$coeff[1]+modelic1$coeff[2]+
    as.matrix(bs.data[c("x1","x2","x3", "x4","x5","x6","x7","x8","x9","x10")])
                                    %*%as.vector(modelic1$coeff[c(-1,-2)])
    bs.data$cont1ic1<- modelic1$coeff[1]+
as.matrix(bs.data[c("x1","x2","x3", "x4","x5","x6","x7","x8","x9","x10")])%*%as.vector(modelic1$coeff[c(-1,-2)])
    ATEDRIC.bs[b]<- mean((bs.data$treat*bs.data$y - (bs.data$treat-bs.data$psbs)*bs.data$treat1ic1)/bs.data$psbs)
        - mean(((1-bs.data$treat)*bs.data$y+(bs.data$treat-bs.data$psbs)*bs.data$cont1ic1)/(1-bs.data$psbs))

  }


  SE.ATT.ipw.n<- sd(ATT.ipw.n.bs)
  SE.ATT.ipw<- sd(ATT.ipw.bs)
  SE.ATE.ipw.n<-sd(ATE.ipw.n.bs)
  SE.ATE.ipw  <-sd(ATE.ipw.bs)
  SE.ATEDRC <- sd(ATEDRC.bs)
  SE.ATEDRIC<- sd(ATEDRIC.bs)

  return(list(SE.ATT.ipw.n,SE.ATT.ipw, SE.ATE.ipw.n, SE.ATE.ipw,SE.ATEDRC,SE.ATEDRIC))
}

##########################################################################
##########################################################################
## Fixing beta for the T.OR
##########################################################################
##########################################################################
beta0<- 0
beta1<- 2
beta2<- 3
beta3<- 2
beta4<- -4
```

```
TATE<- 2 ## True ATE
SNR<- 50 ## Signal noise ratio
set.seed(9999)## Setting the seed
n<- 10 ## Total number of simulations

#############################################################################
#############################################################################
## Defining the variables to store the values of ATT, ATE, SE for ATT,
## SE for ATE by using the different methods Weighted. ASMD for ATT using F.PS,
##Weighted ASMD for ATE using F.PS, Weighted ASMD for ATE using T.PS,
## Weighted ASMD for ATT using T.PS, ASMD for stratification using F.PS,
## ASMD for stratification using T.PS
#############################################################################
#############################################################################


TATT<-ATT.cov<-ATT.reg<-ATT.regc<-ATT.strat<-ATT.stratc<-rep(NA,n)
ATT.ipw.ic<-ATT.ipw.n.ic<-ATT.ipw.c<-ATT.ipw.n.c<-rep(NA,n)
SE.ATT.cov<-SE.ATT.reg<-SE.ATT.regc<-SE.ATT.strat<-SE.ATT.stratc<-rep(NA,n)

TATE<-ATE.cov<-ATE.reg<-ATE.regc<-ATE.strat<-ATE.stratc<-rep(NA,n)
ATE.ipw.n.ic<-ATE.ipw.ic<-ATE.ipw.n.c<-ATE.ipw.c<- rep(NA,n)
ATEDRCOR<-ATEDRCPS<-ATEDRIC<-rep(NA,n)
SE.ATE.cov<- SE.ATE.reg<-SE.ATE.regc<-rep(NA,n)
SE.ATE.strat<-SE.ATE.stratc<-rep(NA,n)

SE.ATT.ipw.ic<-SE.ATT.ipw.ic.n<-SE.ATT.ipw.c<-rep(NA,n)
SE.ATT.ipw.n.c<-SE.ATE.ipw.n.ic<-SE.ATE.ipw.ic<-rep(NA,n)
SE.ATE.ipw.n.c<-SE.ATE.ipw.c<-rep(NA,n)
SEDRCOR<-SEDRCPS<-SEDRIC<-rep(NA,n)

ASMDATEB<-ASMDATE<- ASMDATT<-rep(NA,n)
ASMDATEC<-ASMDATTC<-ASMDATTS<- rep(NA,n)
ASMDATES<- ASMDATTSC<-ASMDATESC<-matrix(NA,n,10)

#############################################################################

##sim=1
for(sim in 1:n){
N<- 1000  ## Sample Size

#############################################################################
#############################################################################
## Generating the dataset for normal distribution
## with mean zero and variance 1
#############################################################################
#############################################################################
x1<- rnorm(N,0,1)
x2<- rnorm(N,0,1)
x3<- rnorm(N,0,1)
x4<- rnorm(N,0,1)
x5<- rnorm(N,0,1)
x6<- rnorm(N,0,1)
x7<- rnorm(N,0,1)
x8<- rnorm(N,0,1)
x9<- rnorm(N,0,1)
x10<- rnorm(N,0,1)
x1.sq<-(x1)^2
x2.sq<-(x2)^2
x.23<- x2*x3
x.345<- x3*x4*x5

#####################################################################
#####################################################################
## betas for treatment selection model
#####################################################################
#####################################################################

beta.low <- log(1.25)
beta.med <- log(1.5)
beta.high <- log(1.75)
beta.v.high <- log(2)
beta.0.treat<- -1.396

#################################################################
#################################################################
## Generating the treatment variable
#################################################################
#################################################################

logit.treat<- beta.0.treat+beta.low*x1 + beta.med*x2 + beta.high*x3
                          + beta.low*x1.sq + beta.med*x2.sq
                                  + beta.high*x.23+ beta.v.high*x.345
p.treat <- exp(logit.treat)/(1 + exp(logit.treat))
treat <- rbinom(N,1,p.treat)
sum(treat)

#################################################################
## Estimating the outcome variable
#################################################################
```

```
#################################################################
sigma<- sqrt(var(ey)/SNR)
y<- ey + rnorm(N, 0, sigma)


## Now creating a dataset
datasim<- data.frame(y,treat,x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x1.sq, x2.sq, x.23, x.345)
#################################################################
#################################################################
## Creating the potential outcomes for each subject
#################################################################
#################################################################
datasim$y1<-beta0+ beta1*x1.sq + beta2*x4
datasim$y0<-  beta0 + beta3*x1*x4+ beta4*x5
##TATE[sim]<- mean(datasim$y1 - datasim$y0)
##TATT[sim]<- mean(datasim$y1[treat==1] - datasim$y0[treat==1])
#################################################################
#################################################################
## Estimate the Propensity score using GBM
#################################################################
#################################################################

model1<- gbm(treat~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10 ,data = datasim,
     distribution = "bernoulli",
     var.monotone = NULL,
     n.trees = 1000,
     interaction.depth = 1,
     n.minobsinnode = 10,
     shrinkage = 0.05,
     bag.fraction = 1.0,
     train.fraction = 1.0,
     cv.folds=10,
     keep.data = TRUE,
     verbose = FALSE,
     n.cores = NULL)
ps1<-predict(model1, datasim, type="response")
likely1<- sum(datasim$treat*log(ps1))+ sum((1-datasim$treat)*log(1-ps1))


model2<- gbm(treat~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10,data = datasim,
     distribution = "bernoulli",
     var.monotone = NULL,
     n.trees = 1000,
     interaction.depth = 2,
     n.minobsinnode = 10,
     shrinkage = 0.05,
     bag.fraction = 1.0,
     train.fraction = 1.0,
     cv.folds=10,
     keep.data = TRUE,
     verbose = FALSE,
     n.cores = NULL)
ps2<-predict(model2, datasim, type="response")
likely2<- sum(datasim$treat*log(ps2))+ sum((1-datasim$treat)*log(1-ps2))

model3<- gbm(treat~ x1+x2+x3+x4+x5+x6+x7+x8+x9+x10,data = datasim,
     distribution = "bernoulli",
     var.monotone = NULL,
     n.trees = 1000,
     interaction.depth = 3,
     n.minobsinnode = 10,
     shrinkage = 0.05,
     bag.fraction = 1.0,
     train.fraction = 1.0,
     cv.folds=10,
     keep.data = TRUE,
     verbose = FALSE,
     n.cores = NULL)
ps3<-predict(model3, datasim, type="response")
likely3<- sum(datasim$treat*log(ps3))+ sum((1-datasim$treat)*log(1-ps3))
#####################################################
#####################################################
## Choosing the best model to estimate GBM
#####################################################
#####################################################
maxlikly<- max(likely1,likely2,likely3)

if(maxlikly==likely1) {datasim$ps=ps1; modelps=model1; depth=1}
if(maxlikly==likely2) {datasim$ps=ps2; modelps=model2; depth=2}
if(maxlikly==likely3) {datasim$ps=ps3; modelps=model3; depth=3}

#################################################################
#################################################################
## Assessing the covariate balance using propensity score as weights
#################################################################
#################################################################

## Estimating the covariate balance using the ASMD.
## Covariate balance when estimating ATE
dataX<- subset(datasim,select=c(3:12))
datagroup1<- subset(datasim, datasim$treat== 1,select =c(3:12,19))
```

```
datagroup0<- subset(datasim, datasim$treat== 0,select =c(3:12,19))

#############################################################
#############################################################
## Estimating ASMD when estimate the ATE
#############################################################
#############################################################
xbar1ATE<- apply(datagroup1,2,function(a,b)weighted.mean(a,b),b=1/datagroup1$ps);
                  xbar1ATE<- t(as.matrix(unname(xbar1ATE[-11])));
xbar0ATE<- apply(datagroup0,2,function(a,b)weighted.mean(a,b),b=1/(1-datagroup0$ps));
                  xbar0ATE<- t(as.matrix(unname(xbar0ATE[-11])));
sdATE<-t(as.matrix(unname(apply(dataX, 2, sd))))
ASMDATE[sim,]<- abs(xbar1ATE-xbar0ATE)/sdATE


#############################################################
#############################################################
## Estimating ASMD when estimate the ATT
#############################################################
#############################################################
xbar1ATT<- apply(datagroup1,2,function(a,b)weighted.mean(a,b),b=rep(1,length(datagroup1$x1)));
                  xbar1ATT<- t(as.matrix(unname(xbar1ATT[-11])));
xbar0ATT<- apply(datagroup0,2,function(a,b)weighted.mean(a,b),b=(datagroup0$ps)/(1-datagroup0$ps));
                        xbar0ATT<- t(as.matrix(unname(xbar0ATT[-11])));
sdATT<-apply(datagroup1, 2, sd); sdATT<-sdATT[-11];
ASMDATT[sim,]<- abs(xbar1ATT-xbar0ATT)/sdATT

#############################################################
#############################################################
##Estimating the ATT using regression method
#############################################################
#############################################################
modelglm<- lm(y~ps*treat, data= datasim)

#############################################################
#############################################################
## Estimating the Variance for ATT When estimated using PS as covariate
#############################################################
varcov<- vcov(modelglm)
cov<- matrix(c(varcov[3,3], varcov[3,4], varcov[4,3], varcov[4,4]),nrow=2, ncol=2)
##Xtreatmean<- matrix(c(1, pstreatmean), nrow=1, ncol=2)
##SE.ATT.cov[sim]<- sqrt(Xtreatmean%*%cov%*%t(Xtreatmean))

#############################################################
### Estimation of ATT using the mean of PS for treatment group and using
### that PS in estimating the ATT
#############################################################

ATT.reg[sim]<-modelglm$coef[3]+modelglm$coef[4]*mean(datasim$ps[datasim$treat==1])
#############################################################

#############################################################
### Estimating the standard error for ATT when mean of PS for treated group
### is used
#############################################################

Xtreatmean1<- matrix(c(1, mean(datasim$ps[datasim$treat==1])), nrow=1, ncol=2)
SE.ATT.reg[sim]<- sqrt(Xtreatmean1%*%cov%*%t(Xtreatmean1))

#############################################################
### Estimating the ATT using startification
#############################################################
breakvals<- as.numeric(quantile(datasim$ps, c(0.40,0.60,0.80)))
datasim$strata1<-cut(datasim$ps,br=c(0,breakvals,1), right=FALSE, labels=c(1:4))
a<- table(datasim$strata1)
a1<- table(datasim$strata1[treat==1])
fits <- lmList( y ~ treat | strata1, data=datasim)
## Gives the ATT via stratification
ATT.strat[sim]<-sum(a1/sum(a1)*summary(fits)$coefficients[,1,2])

#############################################################
### Estimating the SE for ATT stratification
#############################################################
SE.ATT.strat[sim]<-sqrt(sum((a1/sum(a1))^2*(summary(fits)$coefficients[,2,2])^2))
#############################################################
#############################################################
## Estimate the covariate balance after stratification when the PS is estimated using GBM
#############################################################
#############################################################

datat1<- subset(datasim, select=c(3:12,20), datasim$treat== 1)
datat0<- subset(datasim, select=c(3:12,20), datasim$treat== 0)
meant1s<-aggregate(datat1[1:10], by=list(datat1$strata1),FUN=mean)[,-1]
meant0s<-aggregate(datat0[1:10], by=list(datat0$strata1),FUN=mean)[,-1]
meant1sn<-as.matrix(meant1s)
meant0sn<-as.matrix(meant0s)
dms<-abs(meant1sn-meant0sn)
dmsm<-apply(dms,2,mean)
sdtrt<- apply(datat1,2,sd)[-11]

ASMDATTS[sim,]<-dmsm/sdtrt     ## Will provide ASMD estimate when ATT is estimated using stratification
```

```
ASMDATES[sim,]<-dmsm/sdATE       ## Will provide ASMD estimate when ATE is estimated using stratification

#####################################################################
##  Estimating ATT using IPW
#####################################################################
temp.odd<-(1-datasim$treat)*datasim$ps/(1-datasim$ps)

ATT.ipw[sim]<-mean(datasim$y[datasim$treat==1])- mean(datasim$y*temp.odd)        ## Will estimate ATT using IPW

## Will estimate ATT using IPWN
ATT.ipw.n[sim]<-mean(datasim$y[datasim$treat==1])- sum(datasim$y*temp.odd)/sum(temp.odd)

#####################################################################
##  Estimating regression method by using the regression method.
## using the mean of x1, ...., x10
#####################################################################
modelglm<- lm(y~ps*treat, data= datasim)

#############################################################################
### Estimation of ATE using the mean of PS for entire sample and using
### that PS in estimating the ATE
#############################################################################

ATE.reg[sim]<-modelglm$coef[3]+modelglm$coef[4]*mean(datasim$ps)

#############################################################################################
### Estimating the se for ATE when the mean of ps is used
#############################################################################################
Xpopmean1<- matrix(c(1, mean(datasim$ps)), nrow=1, ncol=2)
SE.ATE.reg[sim]<- sqrt(Xpopmean1%*%cov%*%t(Xpopmean1))

#############################################################################
##### Estimating ATE sing the stratification
#############################################################################

ATE.strat[sim]<-sum(a/N*summary(fits)$coefficients[,1,2])##  Gives the ATE

#############################################################################################
## Finding the SE for ATE stratification
#############################################################################################
SE.ATE.strat[sim]<-sqrt(sum((a/N)^2*(summary(fits)$coefficients[,2,2])^2))

#################################################################
##  Estimating ATE using IPW and IPWN
#################################################################

treatnum<- datasim$treat*datasim$y/datasim$ps
treatdenom<- datasim$treat/datasim$ps
contnum<- (1-datasim$treat)*datasim$y/(1-datasim$ps)
contdenom<- (1-datasim$treat)/(1-datasim$ps)

ATE.ipw.n[sim]<- sum(treatnum)/sum(treatdenom) - sum(contnum)/sum(contdenom)
ATE.ipw[sim]<- mean(treatnum)-mean(contnum)
#################################################################33
## Now estimating the ATE using the DR with correct OR
datasim$x1.sq.t<- x1.sq*treat
datasim$x4.t<- x4*treat
datasim$x1x4t<- x1*x4*(1-treat)
datasim$x1x4<- x1*x4
datasim$x5t<- x5*(1-treat)

## Finding the correct multiple regression and finding the predicted values
modelc<- lm(y~x1.sq.t+x4.t+x1x4t+x5t, data = datasim)
## Now finding the predicting values

datasim$treat1c<- modelc$coeff[1]+ as.matrix(datasim[c("x1.sq", "x4")])%*%as.vector(modelc$coeff[c(2,3)])
datasim$cont1c<- modelc$coeff[1]+ as.matrix(datasim[c("x1x4", "x5")])%*%as.vector(modelc$coeff[c(4,5)])

ATEDRC[sim]<- mean((datasim$treat*datasim$y - (datasim$treat-datasim$ps)*datasim$treat1c)/datasim$ps)-
          mean(((1-datasim$treat)*datasim$y+(datasim$treat-datasim$ps)*datasim$cont1c)/(1-datasim$ps))

#############################################################################
##### Now estimating the ATE using the DR with Incorrect OR
#############################################################################
modelic<- lm(y~treat+x1+x2+x3+x4+x5+x6+x7+x8+x9+x10, data=datasim)
datasim$treat1ic<-modelic$coeff[1]+modelic$coeff[2]+
      as.matrix(datasim[c("x1","x2","x3", "x4","x5","x6","x7","x8","x9","x10")])
                          %*%as.vector(modelic$coeff[c(-1,-2)])
datasim$cont1ic<- modelic$coeff[1]+
as.matrix(datasim[c("x1","x2","x3", "x4","x5","x6","x7","x8","x9","x10")])%*%as.vector(modelic$coeff[c(-1,-2)])
ATEDRIC[sim]<- mean((datasim$treat*datasim$y - (datasim$treat-datasim$ps)*datasim$treat1ic)/datasim$ps)-
      mean(((1-datasim$treat)*datasim$y+(datasim$treat-datasim$ps)*datasim$cont1ic)/(1-datasim$ps))

#############################################################################

datasim1<- data.frame(y,treat,x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x1.sq, x2.sq, x.23, x.345)
SE.result<-Var.bootstrap(dataset=datasim1, depth,  nB=200) ## Estimating the SE for IPW and IPWN
SE.ATT.ipw.n[sim]<-SE.result[[1]]
SE.ATT.ipw[sim]<-SE.result[[2]]
SE.ATE.ipw.n[sim]<-SE.result[[3]]
SE.ATE.ipw[sim]<-SE.result[[4]]
```

```
SE.ATEDRC[sim]<- SE.result[[5]]
SE.ATEDRIC[sim]<-SE.result[[6]]

print(sim)

}

#############################################################################
#############################################################################
## Estimating the average of estimates, bias, standard errors, ESE and RMSE

ATT.REG<- mean(ATT.reg); BIAS.ATT.REG<-ATT.REG -ATATT;
ATT.STRAT<- mean(ATT.strat); BIAS.ATT.STRAT<-ATT.STRAT -ATATT;
ATT.IPW<- mean(ATT.ipw);BIAS.ATT.IPW<- ATT.IPW - ATATT;
ATT.IPWN<-mean(ATT.ipw.n); BIAS.ATT.IPWN<- ATT.IPWN - ATATT;


SE.ATT.REG<-mean(SE.ATT.reg)
SE.ATT.STRAT<-mean(SE.ATT.strat)
SE.ATT.IPW<- mean(SE.ATT.ipw)
SE.ATT.IPWN<- mean(SE.ATT.ipw.n)

## Estimating the ESE for ATT
ESE.ATT.REG<-sd(ATT.reg)
ESE.ATT.STRAT<- sd(ATT.strat)
ESE.ATT.IPW<- sd(ATT.ipw)
ESE.ATT.IPWN<- sd(ATT.ipw.n)

### Estimating the RMSE

RMSE.ATT.REG<- sqrt(mean((ATT.reg-ATATT)^2))
RMSE.ATT.STRAT<- sqrt(mean((ATT.strat-ATATT)^2))
RMSE.ATT.IPW<- sqrt(mean((ATT.ipw-ATATT)^2))
RMSE.ATT.IPWN<- sqrt(mean((cTT.ipw.n -ATATT)^2))

## The ATE for cov, strat, ipw, ipwn, drc, dric
ATATE<-1
##ATE.COV<-mean(ATE.cov)
ATE.REG<- mean(ATE.reg);          BIAS.ATE.REG<-ATE.REG -ATATE;
ATE.STRAT<-mean(ATE.strat);       BIAS.ATE.STRAT<-ATE.STRAT -ATATE;
ATE.IPW<-mean(ATE.ipw);           BIAS.ATE.IPW<-ATE.IPW -ATATE;
ATE.IPW.N<-mean(ATE.ipw.n);       BIAS.ATE.IPW.N<-ATE.IPW.N -ATATE;
ATE.DRC<-mean(ATEDRC);            BIAS.ATE.DRC<-ATE.DRC -ATATE;
ATE.DRIC<-mean(ATEDRIC);          BIAS.ATE.DRIC<-ATE.DRIC -ATATE;

## The SE of each method for ATE

SE.ATE.REG<- mean(SE.ATE.reg)
SE.ATE.STRAT<- mean(SE.ATE.strat)
SE.ATE.IPW<-mean(SE.ATE.ipw)
SE.ATE.IPWN<-mean(SE.ATE.ipw.n)
SE.ATEDRC<- mean(SE.ATEDRC)
SE.ATEDRIC<- mean(SE.ATEDRIC)

## Estimating the ESE for ATE

ESE.ATE.REG<- sd(ATE.reg)
ESE.ATE.STRAT<- sd(ATE.strat)
ESE.ATE.IPW<- sd(ATE.ipw)
ESE.ATE.IPWN<- sd(ATE.ipw.n)
ESE.ATE.DRC<- sd(ATEDRC)
ESE.ATE.DRIC<- sd(ATEDRIC)


RMSE.ATE.REG<- sqrt(mean((ATE.reg-ATATE)^2))
RMSE.ATE.STRAT<- sqrt(mean((ATE.strat-ATATE)^2))
RMSE.ATE.IPW<- sqrt(mean((ATE.ipw-ATATE)^2))
RMSE.ATE.IPWN<- sqrt(mean((ATE.ipw.n-ATATE)^2))
RMSE.ATE.DRC<- sqrt(mean((ATEDRC-ATATE)^2))
RMSE.ATE.DRIC<- sqrt(mean((ATEDRIC-ATATE)^2))

#############################################################################
#############################################################################
## R Code producing results for chapter 2 when the propensity score
##is estimated using the GBM and variables are
## dependently normally distributed.
#############################################################################
#############################################################################

## The R Code for dependently normally distributed is same as of independent variables.
## We just need to add an extra package library(mvtnorm)
and the data is to be generated as following

## Creating Matrix for mean and variance covariance Matrix
meanmat<- rep(0,10)
m<- diag(1,10,10)
m[lower.tri(m)] <- 0.50
m[upper.tri(m)] <- 0.50

x <- rmvnorm(n=N, mean=meanmat, sigma=m)
x1<- x[,1]
```

```
x2<- x[,2]
x3<- x[,3]
x4<- x[,4]
x5<- x[,5]
x6<- x[,6]
x7<- x[,7]
x8<- x[,8]
x9<- x[,9]
x10<- x[,10]
x1.sq<- (x1)^2
x2.sq<- (x2)^2
x.23<- x2*x3
x.345<-x3*x4*x5

## The results for ATT and ATE can be estimated by using the code given above.
```

R Code for Chapter 3

```
#############################################################################################
#############################################################################################
## R Code for estimating Generalized propensity score
## using the multinomial logistic regression
#############################################################################################
#############################################################################################

#############################################################################################
#############################################################################################
## Variance function using bootstrap
#############################################################################################
#############################################################################################

## Function for bootstrap
Var.bootstrap<- function(dataset,  nB=50)
{ ATEIPW21.bs<- ATEIPW31.bs<-ATEIPW32.bs<-ATEIPWC21.bs<-ATEIPWC32.bs<-ATEIPWC31.bs<-rep(NA,nB)
ATESTRAT21.bs<-ATESTRAT31.bs<-ATESTRAT32.bs<- rep(NA,nB)
ATESTRATC21.bs<-ATESTRATC31.bs<-ATESTRATC32.bs<- rep(NA,nB)
ATEDR21.bs<-ATEDR31.bs<-ATEDR32.bs<- rep(NA,nB)
ATEDR21co.bs<- ATEDR31co.bs<- ATEDR32co.bs <- rep(NA,nB)
ATEDR21cop.bs<-ATEDR31cop.bs<-ATEDR32cop.bs<- rep(NA,nB)
ATEDR21cp.bs<-ATEDR31cp.bs<-ATEDR32cp.bs<- rep(NA,nB)

  N<-nrow(dataset)
  for(b in 1:nB){
  bs.data<-as.data.frame(dataset[sample(nrow(dataset),N, replace=TRUE),])


#############################################################################################
##Estimating the incorrect PS
#############################################################################################
gpsmn.bs<- multinom(treat~x1+x2+x3+x4+x5+x6, data = bs.data)
psmn.bs<-predict(gpsmn.bs,bs.data,"prob")


## Now assigning Generalized PS to each subject, based upon there treatment group.
for (i in 1:nrow(bs.data)){
if(bs.data$treat[i]==1){
bs.data$GPSM[i]<- psmn.bs[i,1]
}else if(bs.data$treat[i]==2) {
bs.data$GPSM[i]<- psmn.bs[i,2]
}else{
bs.data$GPSM[i]<- psmn.bs[i,3]
}
bs.data$GPSM1[i]<- psmn.bs[i,1]
bs.data$GPSM2[i]<- psmn.bs[i,2]
bs.data$GPSM3[i]<- psmn.bs[i,3]
}

head(bs.data)
#############################################################################
### Estimating ATE using between two treatment groups using the Inverse probablity

ATEIPW<- function(a,b){
mu1<- sum(a$y/a$GPSM)/sum(1/a$GPSM)
mu2<- sum(b$y/b$GPSM)/sum(1/b$GPSM)
ATE<- mu1-mu2
return(ATE)
}

datat1.bs<- subset(bs.data, bs.data$treat== 1)
datat2.bs<- subset(bs.data, bs.data$treat== 2)
datat3.bs<- subset(bs.data, bs.data$treat== 3)
ATEIPW21.bs[b]<-ATEIPW(datat2.bs,datat1.bs)
ATEIPW31.bs[b]<-ATEIPW(datat3.bs,datat1.bs)
ATEIPW32.bs[b]<-ATEIPW(datat3.bs,datat2.bs)
```

```
##############################################################################
##############################################################################

##############################################################################
##############################################################################
## Estimating ATE using the stratification using the incorrect GPS model

breakvals1.bs<- as.numeric(quantile(bs.data$GPSM1, c(0.20, 0.40,0.60,0.80)))
bs.data$strata1<-cut(bs.data$GPSM1,br=c(0,breakvals1.bs,1), right=FALSE, labels=c(1:5))
a1.bs<- table(bs.data$strata1)
datastratatreat1.bs<- subset(bs.data, bs.data$treat== 1)
MUHAT1.bs<-tapply(datastratatreat1.bs$y, datastratatreat1.bs$strata1,mean)
EY1.bs<- sum(a1.bs/sum(a1.bs)*MUHAT1.bs)


breakvals2.bs<- as.numeric(quantile(bs.data$GPSM2, c(0.20, 0.40,0.60,0.80)))
bs.data$strata2<-cut(bs.data$GPSM2,br=c(0,breakvals2.bs,1), right=FALSE, labels=c(1:5))
a2.bs<- table(bs.data$strata2)
datastratatreat2.bs<- subset(bs.data, bs.data$treat== 2)
MUHAT2.bs<-tapply(datastratatreat2.bs$y, datastratatreat2.bs$strata2,mean)
EY2.bs<- sum(a2.bs/sum(a2.bs)*MUHAT2.bs)


breakvals3.bs<- as.numeric(quantile(bs.data$GPSM3, c(0.20, 0.40,0.60,0.80)))
bs.data$strata3<-cut(bs.data$GPSM3,br=c(0,breakvals3.bs,1), right=FALSE, labels=c(1:5))
a3.bs<- table(bs.data$strata3)
datastratatreat3.bs<- subset(bs.data, datasim$treat== 3)
MUHAT3.bs<-tapply(datastratatreat3.bs$y, datastratatreat3.bs$strata3,mean)
EY3.bs<- sum(a3.bs/sum(a3.bs)*MUHAT3.bs)


ATESTRAT21.bs[b]<- EY2.bs - EY1.bs
ATESTRAT31.bs[b]<- EY3.bs - EY1.bs
ATESTRAT32.bs[b]<- EY3.bs - EY2.bs

##############################################################################
##############################################################################
## Estimatin the ATE using the using the Doubly Robust estimator

## First creating the dummy variables in for treatment 1 and treatment 1
bs.data$treat1<- ifelse(bs.data$treat==1,1,0)
bs.data$treat2<- ifelse(bs.data$treat==2,1,0)
bs.data$treat3<- ifelse(bs.data$treat==3,1,0)

##############################################################################
## DR using the Incorrect Propensity scores and Correct OR

## Estimating the outcome regression of the

modelG1.bs<-lm(y~x1+x2+x3+x4+x5+x6, data=datat1.bs)
modelG2.bs<-lm(y~x1+x2+x3+x4+x5+x6, data=datat2.bs)
modelG3.bs<-lm(y~x1+x2+x3+x4+x5+x6, data=datat3.bs)

##modelmc<- lm(y~treat1+treat2+x1+x2+x3+x4+x5+x6, data=datasim)
bs.data$OR1<- modelG1.bs$coeff[1]+
                 as.matrix(bs.data[c("x1","x2","x3", "x4","x5","x6")])%*%as.vector(modelG1.bs$coeff[c(-1)])
bs.data$OR2<- modelG2.bs$coeff[1]+
                 as.matrix(bs.data[c("x1","x2","x3", "x4","x5","x6")])%*%as.vector(modelG2.bs$coeff[c(-1)])
bs.data$OR3<- modelG3.bs$coeff[1]+
                as.matrix(bs.data[c("x1","x2","x3", "x4","x5","x6")])%*%as.vector(modelG3.bs$coeff[c(-1)])


MUDR1.bs<- mean((bs.data$treat1*bs.data$y - (bs.data$treat1-bs.data$GPSM1)*bs.data$OR1)/bs.data$GPSM1)
MUDR2.bs<- mean((bs.data$treat2*bs.data$y - (bs.data$treat2-bs.data$GPSM2)*bs.data$OR2)/bs.data$GPSM2)
MUDR3.bs<- mean((bs.data$treat3*bs.data$y - (bs.data$treat3-bs.data$GPSM3)*bs.data$OR3)/bs.data$GPSM3)

ATEDR21co.bs[b]<- MUDR2.bs-MUDR1.bs
ATEDR31co.bs[b]<- MUDR3.bs-MUDR1.bs
ATEDR32co.bs[b]<- MUDR3.bs-MUDR2.bs

##############################################################################
##############################################################################
## Estimating the correct outcome regression

modelmc1.bs<- lm(y~x1+x4, data=datat1.bs)
modelmc2.bs<-lm(y~x2+x5, data=datat2.bs)
modelmc3.bs<-lm(y~x3+x6, data=datat3.bs)

bs.data$OR1c<- modelmc1.bs$coeff[1]+
                 as.matrix(bs.data[c("x1","x4")])%*%as.vector(modelmc1.bs$coeff[c(-1)])
bs.data$OR2c<- modelmc2.bs$coeff[1]+
                 as.matrix(bs.data[c("x2","x5")])%*%as.vector(modelmc2.bs$coeff[c(-1)])
bs.data$OR3c<- modelmc3.bs$coeff[1]+
                 as.matrix(bs.data[c("x3","x6")])%*%as.vector(modelmc3.bs$coeff[c(-1)])

## Estimating the treatment effect using the TRUE OR
MUDR1co.bs<- mean((bs.data$treat1*bs.data$y - (bs.data$treat1-bs.data$GPSM1)*bs.data$OR1c)/bs.data$GPSM1)
MUDR2co.bs<- mean((bs.data$treat2*bs.data$y - (bs.data$treat2-bs.data$GPSM2)*bs.data$OR2c)/bs.data$GPSM2)
MUDR3co.bs<- mean((bs.data$treat3*bs.data$y - (bs.data$treat3-bs.data$GPSM3)*bs.data$OR3c)/bs.data$GPSM3)
```

```
ATEDR21.bs[b]<- MUDR2co.bs-MUDR1co.bs
ATEDR31.bs[b]<- MUDR3co.bs-MUDR1co.bs
ATEDR32.bs[b]<- MUDR3co.bs-MUDR2co.bs

#############################################################################################


}
SE.ATEIPW21.bs<-sd(ATEIPW21.bs)
SE.ATEIPW31.bs<-sd(ATEIPW31.bs)
SE.ATEIPW32.bs<-sd(ATEIPW32.bs)

##SE.ATEIPWC21.bs<-sd(ATEIPWC21.bs)
##SE.ATEIPWC31.bs<-sd(ATEIPWC31.bs)
##SE.ATEIPWC32.bs<-sd(ATEIPWC32.bs)

SE.ATESTRAT21.bs<-sd(ATESTRAT21.bs)
SE.ATESTRAT31.bs<-sd(ATESTRAT31.bs)
SE.ATESTRAT32.bs<-sd(ATESTRAT32.bs)

##SE.ATESTRATC21.bs<-sd(ATESTRATC21.bs)
##SE.ATESTRATC31.bs<-sd(ATESTRATC31.bs)
##SE.ATESTRATC32.bs<-sd(ATESTRATC32.bs)

SE.ATEDR21.bs<- sd(ATEDR21.bs)
SE.ATEDR31.bs<- sd(ATEDR31.bs)
SE.ATEDR32.bs<- sd(ATEDR32.bs)

SE.ATEDR21co.bs<-sd(ATEDR21co.bs)
SE.ATEDR31co.bs<-sd(ATEDR31co.bs)
SE.ATEDR32co.bs<-sd(ATEDR32co.bs)

##SE.ATEDR21cop.bs<-sd(ATEDR21cop.bs)
##SE.ATEDR31cop.bs<-sd(ATEDR31cop.bs)
##SE.ATEDR32cop.bs<-sd(ATEDR32cop.bs)

##SE.ATEDR21cp.bs<-sd(ATEDR21cp.bs)
##SE.ATEDR31cp.bs<-sd(ATEDR31cp.bs)
##SE.ATEDR32cp.bs<-sd(ATEDR32cp.bs)



  return(list(SE.ATEIPW21.bs, SE.ATEIPW31.bs, SE.ATEIPW32.bs,SE.ATESTRAT21.bs,SE.ATESTRAT31.bs,SE.ATESTRAT32.bs,
              SE.ATEDR21.bs,SE.ATEDR31.bs,SE.ATEDR32.bs, SE.ATEDR21co.bs, SE.ATEDR31co.bs, SE.ATEDR32co.bs))
}



##install.packages("reshape")
library(mvtnorm)
library(miscF)
library( MCMCpack)
library(coda)
library(MASS)
require(stats)
library(nnet)
library(rpart)
library(reshape)
library(Hmisc)
library(lattice)
library(survival)
library(splines)
library(Formula)
library(ggplot2)
library(Matrix)
##library(lme4)
set.seed(9998)

beta2<- matrix(c(0.25, 0.25, 0.25,-0.25, 0.25, 0.25,0.25,0.25),1,8)
beta3<- matrix(c(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1),1,8)
tau1<-0
tau2<-0.5

n<-40

ASMDATE <- ASMDATT <- KSATE<- PKSATT<-KSATEB<-ASMDATEB<-matrix(NA,n,6)

ATEIPW21<-ATEIPW31<-ATEIPW32<-ATEDR21<-ATEDR31<-ATEDR32<-rep(NA,n)
ATEDR21co<- ATEDR31co<-ATEDR32co<-ATESTRAT21<-ATESTRAT31<-ATESTRAT32<-rep(NA,n)
ATESTRAT21N<-ATESTRAT31N<-ATESTRAT32N<-rep(NA,n)
ATESTRAT21NC<-ATESTRAT31NC<-ATESTRAT32NC<-rep(NA,n)
ATEDR21cp<- ATEDR31cp<-ATEDR32cp<-ATEDR21cop<- ATEDR31cop<-ATEDR32cop<-rep(NA,n)
ATESTRATC21<-ATESTRATC32<-ATESTRATC31<-rep(NA,n)
ATEIPWC21<-ATEIPWC31<-ATEIPWC32<-rep(NA,n)
TATE21<-TATE31<-TATE32<-rep(NA,n)
SE.ATEIPW21.bs<-SE.ATEIPW31.bs<-SE.ATEIPW32.bs<-rep(NA,n)
SE.ATEIPWC21.bs<-SE.ATEIPWC31.bs<- SE.ATEIPWC32.bs<- rep(NA,n)
```

```
SE.ATESTRAT21.bs<-SE.ATESTRAT31.bs<-SE.ATESTRAT32.bs<-rep(NA,n)
SE.ATESTRATC21.bs<-SE.ATESTRATC31.bs<-SE.ATESTRATC32.bs<-rep(NA,n)
SE.ATEDR21.bs<-SE.ATEDR31.bs<-SE.ATEDR32.bs<-rep(NA,n)
SE.ATEDR21co.bs<-SE.ATEDR31co.bs<- SE.ATEDR32co.bs<- rep(NA,n)
SE.ATEDR21cop.bs<-SE.ATEDR31cop.bs<- SE.ATEDR32cop.bs<- rep(NA,n)
SE.ATEDR21cp.bs<-SE.ATEDR31cp.bs<- SE.ATEDR32cp.bs<- rep(NA,n)
SESTRATA21<-SESTRATA31<-SESTRATA32<- rep(NA,n)

for(j in 1:n){
meanmat<- rep(0,3)
m<- matrix(c(2,1,-1,1,1,-0.5,-1,-0.5,1),3,3)
N<-1000
x <- rmvnorm(N, mean=meanmat, sigma=m)

datasim<- data.frame(x[,1],x[,2],x[,3],runif(N,-3,3),rchisq(N, 1, ncp = 0),
                     rbinom(N,1,0.5), (x[,1]^2), (x[,2]^2), (x[,3]*x[,2]) )
names(datasim)<-c("x1","x2","x3", "x4","x5","x6","x1.sq","x2.sq","x.23")
datasim$x.345<- datasim$x3*datasim$x4*datasim$x5

treat2<-exp(as.matrix(datasim[c("x1","x2","x3", "x4","x1.sq","x2.sq","x.23","x.345")])%*%t(beta2))
treat3<-exp(as.matrix(datasim[c("x1","x2","x3", "x4","x1.sq","x2.sq","x.23","x.345")])%*%t(beta3))
denom<- 1+treat2+treat3; p1<- 1/denom; p2<- treat2/denom; p3<- treat3/denom
datasim$treat <- rMultinom(p=cbind(p1, p2, p3),1)
##table(datasim$treat)


## Generating the outcome variable
y1<- datasim$x1+datasim$x2+datasim$x3+datasim$x4+datasim$x5-1+datasim$x6-0.5+rnorm(N,0,1)
y2<- 2*datasim$x1+3*datasim$x2+datasim$x3+2*datasim$x4+2*(datasim$x5-1)+2*(datasim$x6-0.5)+tau1+rnorm(N,0,1)
y3<- 3*datasim$x1+datasim$x2+2*datasim$x3-datasim$x4-(datasim$x5-1)-(datasim$x6-0.5)+ tau2+rnorm(N,0,1)

## The obsereved potential outcome
for (i in 1:N){
        if(datasim$treat[i]==1){
                datasim$y[i]<- y1[i]
        }else if(datasim$treat[i]==2) {
                datasim$y[i]<- y2[i]
        }else{
                datasim$y[i]<- y3[i]
        }
}


#############################################################################
#############################################################################
## Estimating the generalized PS using the multinomial logistic regression
gpsmn<- multinom(treat~x1+x2+x3+x4+x5+x6, data = datasim)
psmn<-predict(gpsmn,datasim,"prob")


## Now assigning Generalized PS to each subject, based upon there treatment group.
for (i in 1:nrow(datasim)){
if(datasim$treat[i]==1){
datasim$GPSM[i]<- psmn[i,1]
}else if(datasim$treat[i]==2) {
datasim$GPSM[i]<- psmn[i,2]
}else{
datasim$GPSM[i]<- psmn[i,3]
}
datasim$GPSM1[i]<- psmn[i,1]
datasim$GPSM2[i]<- psmn[i,2]
datasim$GPSM3[i]<- psmn[i,3]
}
#############################################################################################
#############################################################################################

#############################################################################################
#############################################################################################
### Estimating ATE using between two treatment groups using the Inverse probablity

ATEIPW<- function(a,b){
mu1<- sum(a$y/a$GPSM)/sum(1/a$GPSM)
mu2<- sum(b$y/b$GPSM)/sum(1/b$GPSM)
ATE<- mu1-mu2
return(ATE)
}

datat1<- subset(datasim, datasim$treat== 1)
datat2<- subset(datasim, datasim$treat== 2)
datat3<- subset(datasim, datasim$treat== 3)
ATEIPW21[j]<-ATEIPW(datat2,datat1)
ATEIPW31[j]<-ATEIPW(datat3,datat1)
ATEIPW32[j]<-ATEIPW(datat3,datat2)


#############################################################################################
#############################################################################################
## Estimating ATE using the stratification using the incorrect GPS model
datasim$treat1<- ifelse(datasim$treat==1,1,0)
datasim$treat2<- ifelse(datasim$treat==2,1,0)
datasim$treat3<- ifelse(datasim$treat==3,1,0)
```

```
breakvals1<- as.numeric(quantile(datasim$GPSM1, c(0.20, 0.40,0.60,0.80)))
datasim$strata1<-cut(datasim$GPSM1,br=c(0,breakvals1,1), right=FALSE, labels=c(1:5))
a1<- table(datasim$strata1)
datastratatreat1<- subset(datasim, datasim$treat== 1)
MUHAT1<-tapply(datastratatreat1$y, datastratatreat1$strata1,mean)
EY1<- sum(a1/sum(a1)*MUHAT1)


breakvals2<- as.numeric(quantile(datasim$GPSM2, c(0.20, 0.40,0.60,0.80)))
datasim$strata2<-cut(datasim$GPSM2,br=c(0,breakvals2,1), right=FALSE, labels=c(1:5))
a2<- table(datasim$strata2)
datastratatreat2<- subset(datasim, datasim$treat== 2)
MUHAT2<-tapply(datastratatreat2$y, datastratatreat2$strata2,mean)
EY2<- sum(a2/sum(a2)*MUHAT2)


breakvals3<- as.numeric(quantile(datasim$GPSM3, c(0.20, 0.40,0.60,0.80)))
datasim$strata3<-cut(datasim$GPSM3,br=c(0,breakvals3,1), right=FALSE, labels=c(1:5))
a3<- table(datasim$strata3)
datastratatreat3<- subset(datasim, datasim$treat== 3)
MUHAT3<-tapply(datastratatreat3$y, datastratatreat3$strata3,mean)
EY3<- sum(a3/sum(a3)*MUHAT3)


ATESTRAT21[j]<- EY2 - EY1
ATESTRAT31[j]<- EY3 - EY1
ATESTRAT32[j]<- EY3 - EY2

## Estimating the variances.
VARMUHAT1<-tapply(datastratatreat1$y, datastratatreat1$strata1,var)
VARMUHAT2<-tapply(datastratatreat2$y, datastratatreat2$strata2,var)
VARMUHAT3<-tapply(datastratatreat3$y, datastratatreat3$strata3,var)
a1t<- table(datasim$strata1[datasim$treat==1])
a2t<- table(datasim$strata2[datasim$treat==2])
a3t<- table(datasim$strata3[datasim$treat==3])
VARSTRATA1<- sum((a1t/sum(a1))^2*VARMUHAT1)
VARSTRATA2<- sum((a2t/sum(a2))^2*VARMUHAT2)
VARSTRATA3<- sum((a3t/sum(a3))^2*VARMUHAT3)


SESTRATA21[j]<- sqrt(VARSTRATA2 + VARSTRATA1)
SESTRATA31[j]<- sqrt(VARSTRATA3 + VARSTRATA1)
SESTRATA32[j]<- sqrt(VARSTRATA3 + VARSTRATA2)


##############################################################################
##############################################################################
##############################################################################
##############################################################################

## Estimatin the ATE using the using the Doubly Robust estimator
##############################################################################
## DR using the Incorrect Propensity scores and Correct OR

## Estimating the outcome regression of the
modelG1<-lm(y~x1+x2+x3+x4+x5+x6, data=datat1)
modelG2<-lm(y~x1+x2+x3+x4+x5+x6, data=datat2)
modelG3<-lm(y~x1+x2+x3+x4+x5+x6, data=datat3)

##modelmc<- lm(y~treat1+treat2+x1+x2+x3+x4+x5+x6, data=datasim)
datasim$OR1<- modelG1$coeff[1]+
                        as.matrix(datasim[c("x1","x2","x3", "x4","x5","x6")])%*%as.vector(modelG1$coeff[c(-1)])
datasim$OR2<- modelG2$coeff[1]+
                        as.matrix(datasim[c("x1","x2","x3", "x4","x5","x6")])%*%as.vector(modelG2$coeff[c(-1)])
datasim$OR3<- modelG3$coeff[1]+
                       as.matrix(datasim[c("x1","x2","x3", "x4","x5","x6")])%*%as.vector(modelG3$coeff[c(-1)])


MUDR1<- mean((datasim$treat1*datasim$y - (datasim$treat1-datasim$GPSM1)*datasim$OR1)/datasim$GPSM1)
MUDR2<- mean((datasim$treat2*datasim$y - (datasim$treat2-datasim$GPSM2)*datasim$OR2)/datasim$GPSM2)
MUDR3<- mean((datasim$treat3*datasim$y - (datasim$treat3-datasim$GPSM3)*datasim$OR3)/datasim$GPSM3)

ATEDR21co[j]<- MUDR2-MUDR1
ATEDR31co[j]<- MUDR3-MUDR1
ATEDR32co[j]<- MUDR3-MUDR2

##############################################################################
##############################################################################
## DR using the Incorrect Propensity scores and Inorrect OR
## Estimating the correct outcome regression

modelmc1<- lm(y~x1+x4, data=datat1)
modelmc2<-lm(y~x2+x5, data=datat2)
modelmc3<-lm(y~x3+x6, data=datat3)

datasim$OR1c<- modelmc1$coeff[1]+
                        as.matrix(datasim[c("x1","x4")])%*%as.vector(modelmc1$coeff[c(-1)])
datasim$OR2c<- modelmc2$coeff[1]+
                        as.matrix(datasim[c("x2","x5")])%*%as.vector(modelmc2$coeff[c(-1)])
```

```
datasim$OR3c<- modelmc3$coeff[1]+
                        as.matrix(datasim[c("x3","x6")])%*%as.vector(modelmc3$coeff[c(-1)])

## Estimating the treatment effect using the TRUE OR
MUDR1co<- mean((datasim$treat1*datasim$y - (datasim$treat1-datasim$GPSM1)*datasim$OR1c)/datasim$GPSM1)
MUDR2co<- mean((datasim$treat2*datasim$y - (datasim$treat2-datasim$GPSM2)*datasim$OR2c)/datasim$GPSM2)
MUDR3co<- mean((datasim$treat3*datasim$y - (datasim$treat3-datasim$GPSM3)*datasim$OR3c)/datasim$GPSM3)


ATEDR21[j]<- MUDR2co-MUDR1co
ATEDR31[j]<- MUDR3co-MUDR1co
ATEDR32[j]<- MUDR3co-MUDR2co

#############################################################################
#############################################################################
#############################################################################


## Generating the potential outcomes for the three treatment groups
## Generating the outcome variable
datasim$y1<- datasim$x1+datasim$x2+datasim$x3+datasim$x4+datasim$x5-1+datasim$x6-0.5
datasim$y2<- 2*datasim$x1+3*datasim$x2+datasim$x3+2*datasim$x4+2*(datasim$x5-1)+2*(datasim$x6-0.5)+tau1
datasim$y3<- 3*datasim$x1+datasim$x2+2*datasim$x3-datasim$x4-(datasim$x5-1)-(datasim$x6-0.5)+tau2

TATE21[j]<- mean(datasim$y2 - datasim$y1)
TATE31[j]<- mean(datasim$y3 - datasim$y1)
TATE32[j]<- mean(datasim$y3 - datasim$y2)

#############################################################################
#############################################################################

## Assessing the balance for each covariate using ASMD, when estimating ATE
## Subsetting datasets for each treatment group
dataX<- subset(datasim,select =c(1:6))
datagroupX1<- subset(datasim, datasim$treat==1 ,select =c(1:6,13))
datagroupX2<- subset(datasim, datasim$treat==2 ,select =c(1:6,13))
datagroupX3<- subset(datasim, datasim$treat==3 ,select =c(1:6,13))


## Now estimating the weighted mean of each covariate after estimating the GPS
xbark1<-apply(datagroupX1, 2, function(a,b)weighted.mean(a,b),b=1/datagroupX1$GPSM); xbark1<- xbark1[-7];
xbark2<-apply(datagroupX2, 2, function(a,b)weighted.mean(a,b),b=1/datagroupX2$GPSM); xbark2<- xbark2[-7];
xbark3<-apply(datagroupX3, 2, function(a,b)weighted.mean(a,b),b=1/datagroupX3$GPSM); xbark3<- xbark3[-7];

## Estimating the unweighted mean and variance for each variable
xbarkp<- apply(dataX, 2, mean)
sdkp<- apply(dataX, 2, sd)
ASMD1<- abs(xbark1-xbarkp)/sdkp
ASMD2<- abs(xbark2-xbarkp)/sdkp
ASMD3<- abs(xbark3-xbarkp)/sdkp
ASMDATE[j,]<- apply(rbind(ASMD1,ASMD2,ASMD3),2,mean)

## Now estimating the ASMD of covariates before
xbark1b<-apply(datagroupX1, 2, mean); xbark1b<- xbark1b[-7];
xbark2b<-apply(datagroupX2, 2, mean); xbark2b<- xbark2b[-7];
xbark3b<-apply(datagroupX3, 2, mean); xbark3b<- xbark3b[-7];

ASMD1B<- abs(xbark1b-xbarkp)/sdkp
ASMD2B<- abs(xbark2b-xbarkp)/sdkp
ASMD3B<- abs(xbark3b-xbarkp)/sdkp

ASMDATEB[j,]<- apply(rbind(ASMD1B,ASMD2B,ASMD3B),2,mean)



#############################################################################
#############################################################################
## Assessing the balance using the KS statistic
#############################################################################
### KS statistic for Group
#############################################################################
prop<-table(datasim$x6)
prop0<-prop[1]/N
prop0<-unname(prop0)

datax16<- head(datagroupX1)[6:7]
datasx16<-datax16[order(datax16[1]),]
datasx16$ecdf<- cumsum(1/datasx16$GPSM/sum(1/datasx16$GPSM))
datasx16$ecdfb<-1/nrow(datasx16)
datasx16$ecdfbn<-cumsum(datasx16$ecdfb)
datasx160<- subset(datasx16, datasx16$x6==0)
ladd<-dim(datasx160)
cumx160<-datasx160[ladd[1], 3]
ksx16<-abs(cumx160-prop0)
cumx160b<-datasx160[ladd[1], 5]
ksx16b<-abs(cumx160b-prop0)


datax26<- datagroupX2[6:7]
datasx26<-datax26[order(datax26[1]),]
datasx26$ecdf<- cumsum(1/datasx26$GPSM/sum(1/datasx26$GPSM))
```

```
datasx26$ecdfb<-1/nrow(datasx26)
datasx26$ecdfbn<-cumsum(datasx26$ecdfb)
datasx260<- subset(datasx26, datasx26$x6==0)
ladd<-dim(datasx260)
cumx260<-datasx260[ladd[1], 3]
ksx26<-abs(cumx260-prop0)
cumx260b<-datasx260[ladd[1], 5]
ksx26b<-abs(cumx260b-prop0)

datax36<- datagroupX3[6:7]
datasx36<-datax36[order(datax36[1]),]
datasx36$ecdf<- cumsum(1/datasx36$GPSM/sum(1/datasx36$GPSM))
datasx36$ecdfb<-1/nrow(datasx36)
datasx36$ecdfbn<-cumsum(datasx36$ecdfb)
datasx360<- subset(datasx36, datasx36$x6==0)
ladd<-dim(datasx360)
cumx360<-datasx360[ladd[1], 3]
ksx36<-abs(cumx360-prop0)
cumx360b<-datasx360[ladd[1], 5]
ksx36b<-abs(cumx360b-prop0)



KS1<- KS2<-KS3<-rep(NA,5)
for(i in 1:5){
datasort1<-datagroupX1[order(datagroupX1[i]),]
datasort1$ecdf<- cumsum(1/datasort1$GPSM/sum(1/datasort1$GPSM))
datasort2<-datagroupX2[order(datagroupX2[i]),]
datasort2$ecdf<- cumsum(1/datasort2$GPSM/sum(1/datasort2$GPSM))
datasort3<-datagroupX3[order(datagroupX3[i]),]
datasort3$ecdf<- cumsum(1/datasort3$GPSM/sum(1/datasort3$GPSM))
t<-sort(c(datasort1[,i],datasort2[,i],datasort3[,i] ))
edfp<- cumsum(rep(1/N,length(t)))
p1<-p2<-p3<-numeric(length(t))
for(k in 1: length(t)){
t0<-t[k]
l1<-sum(datasort1[,i]<=t0)
l2<-sum(datasort2[,i]<=t0)
l3<-sum(datasort3[,i]<=t0)
if(l1==0){
p1[k]<-0
}else {
p1[k]<-datasort1[l1,8]
D1<-cbind(t,abs(p1-edfp))
KS1[i]<-max(D1[,2])
}
if(l2==0){
p2[k]<-0
}else {
p2[k]<-datasort2[l2,8]
D2<-cbind(t,abs(p2-edfp))
KS2[i]<-max(D2[,2])
}
if(l3==0){
p3[k]<-0
}else {
p3[k]<-datasort3[l3,8]
D3<-cbind(t,abs(p3-edfp))
KS3[i]<-max(D3[,2])
}
}
}

PKS1<-as.matrix(KS1)
PKS1<-unname(t(rbind(PKS1, ksx16)))

PKS2<-as.matrix(KS2)
PKS2<-unname(t(rbind(PKS2, ksx26)))

PKS3<-as.matrix(KS3)
PKS3<-unname(t(rbind(PKS3, ksx36)))

KSATE[j,]<- apply(rbind(PKS1,PKS2,PKS3),2,mean)
#############################################################################################
##KS Statistic before adjusting for PS

KSB1<- KSB2<-KSB3<-rep(NA,5)
for(i in 1:5){
datasort1<-datagroupX1[order(datagroupX1[i]),]
datasort1$ecdfb<-1/nrow(datasort1)
datasort1$ecdfbn<- cumsum(datasort1$ecdfb)
datasort2<-datagroupX2[order(datagroupX2[i]),]
datasort2$ecdfb<-1/nrow(datasort2)
datasort2$ecdfbn<- cumsum(datasort2$ecdfb)
datasort3<-datagroupX3[order(datagroupX3[i]),]
datasort3$ecdfb<-1/nrow(datasort3)
datasort3$ecdfbn<- cumsum(datasort3$ecdfb)
t<-sort(c(datasort1[,i],datasort2[,i],datasort3[,i] ))
edfp<- cumsum(rep(1/N,length(t)))
p1<-p2<-p3<-numeric(length(t))
for(k in 1: length(t)){
```

```
t0<-t[k]
l1<-sum(datasort1[,i]<=t0)
l2<-sum(datasort2[,i]<=t0)
l3<-sum(datasort3[,i]<=t0)
if(l1==0){
p1[k]<-0
}else {
p1[k]<-datasort1[l1,9]
D1<-cbind(t,abs(p1-edfp))
KSB1[i]<-max(D1[,2])
}
if(l2==0){
p2[k]<-0
}else {
p2[k]<-datasort2[l2,9]
D2<-cbind(t,abs(p2-edfp))
KSB2[i]<-max(D2[,2])
}
if(l3==0){
p3[k]<-0
}else {
p3[k]<-datasort3[l3,9]
D3<-cbind(t,abs(p3-edfp))
KSB3[i]<-max(D3[,2])
}
}
}

PKSB1<-as.matrix(KSB1)
PKSB1<-unname(t(rbind(PKSB1, ksx16b)))

PKSB2<-as.matrix(KSB2)
PKSB2<-unname(t(rbind(PKSB2, ksx26b)))

PKSB3<-as.matrix(KSB3)
PKSB3<-unname(t(rbind(PKSB3, ksx36b)))

KSATEB[j,]<- apply(rbind(PKSB1,PKSB2,PKSB3),2,mean)



datasim1<- datasim[,1:12]
SE.result<-Var.bootstrap(dataset=datasim1,  nB=50)



SE.ATEIPW21.bs[j]<- SE.result[[1]]
SE.ATEIPW31.bs[j]<- SE.result[[2]]
SE.ATEIPW32.bs[j]<- SE.result[[3]]
SE.ATESTRAT21.bs[j]<-SE.result[[4]]
SE.ATESTRAT31.bs[j]<-SE.result[[5]]
SE.ATESTRAT32.bs[j]<-SE.result[[6]]
SE.ATEDR21.bs[j]<-SE.result[[7]]
SE.ATEDR31.bs[j]<-SE.result[[8]]
SE.ATEDR32.bs[j]<-SE.result[[9]]
SE.ATEDR21co.bs[j]<-SE.result[[10]]
SE.ATEDR31co.bs[j]<- SE.result[[11]]
SE.ATEDR32co.bs[j]<-SE.result[[12]]

print(j)
}

#############################################################################
#############################################################################
TTATE21<-mean(TATE21)
TTATE31<-mean(TATE31)
TTATE32<-mean(TATE32)

MATEIPW21<-mean(ATEIPW21)
MATEIPW31<-mean(ATEIPW31)
MATEIPW32<-mean(ATEIPW32)

SATE21<-mean(ATESTRAT21)
SATE31<-mean(ATESTRAT31)
SATE32<-mean(ATESTRAT32)



MATEDR21<-mean(ATEDR21)
MATEDR31<-mean(ATEDR31)
MATEDR32<-mean(ATEDR32)

MATEDRCO21<-mean(ATEDR21co)
MATEDRCO31<-mean(ATEDR31co)
MATEDRCO32<-mean(ATEDR32co)


#############################################################################
#############################################################################
## Standard Error of the
```

```
SE.IPW21<-mean(SE.ATEIPW21.bs)
SE.IPW31<-mean(SE.ATEIPW31.bs)
SE.IPW32<-mean(SE.ATEIPW32.bs)


SE.STRAT21<-mean(SE.ATESTRAT21.bs)
SE.STRAT31<-mean(SE.ATESTRAT31.bs)
SE.STRAT32<-mean(SE.ATESTRAT32.bs)


SE.DR21<-mean(SE.ATEDR21.bs)
SE.DR31<-mean(SE.ATEDR31.bs)
SE.DR32<-mean(SE.ATEDR32.bs)

SE.DRCO21<-mean(SE.ATEDR21co.bs)
SE.DRCO31<-mean(SE.ATEDR31co.bs)
SE.DRCO32<-mean(SE.ATEDR32co.bs)

#############################################################################
#############################################################################
## EStimated Standard Error

SATEIPW21<-sd(ATEIPW21)
SATEIPW31<-sd(ATEIPW31)
SATEIPW32<-sd(ATEIPW32)


SSATE21<-sd(ATESTRAT21)
SSATE31<-sd(ATESTRAT31)
SSATE32<-sd(ATESTRAT32)


SATEDR21<-sd(ATEDR21)
SATEDR31<-sd(ATEDR31)
SATEDR32<-sd(ATEDR32)

SATEDRCO21<-sd(ATEDR21co)
SATEDRCO31<-sd(ATEDR31co)
SATEDRCO32<-sd(ATEDR32co)


TTATE<-0

RMSE.IPW21<- sqrt(mean((ATEIPW21-tau1)^2))
RMSE.IPW31<- sqrt(mean((ATEIPW31-tau2)^2))
RMSE.IPW32<- sqrt(mean((ATEIPW32-tau1)^2))

RMSE.STRATA21<-sqrt(mean((ATESTRAT21-tau1)^2))
RMSE.STRATA31<-sqrt(mean((ATESTRAT31-tau2)^2))
RMSE.STRATA32<-sqrt(mean((ATESTRAT32-tau1)^2))

RMSE.DR21<-sqrt(mean((ATEDR21-tau1)^2))
RMSE.DR31<-sqrt(mean((ATEDR31-tau2)^2))
RMSE.DR32<-sqrt(mean((ATEDR32-tau1)^2))

RMSE.DRCO21<-sqrt(mean((ATEDR21co-tau1)^2))
RMSE.DRCO31<-sqrt(mean((ATEDR31co-tau2)^2))
RMSE.DRCO32<-sqrt(mean((ATEDR32co-tau1)^2))


#############################################################################

LCLIPW21<-ATEIPW21-1.96*SE.ATEIPW21.bs
UCLIPW21<-ATEIPW21+1.96*SE.ATEIPW21.bs
mean(ifelse(LCLIPW21<tau1&UCLIPW21>tau1, 1,0))

LCLIPW31<-ATEIPW31-1.96*SE.ATEIPW31.bs
UCLIPW31<-ATEIPW31+1.96*SE.ATEIPW31.bs
mean(ifelse(LCLIPW31<tau2&UCLIPW31>tau2, 1,0))

LCLIPW32<-ATEIPW32-1.96*SE.ATEIPW32.bs
UCLIPW32<-ATEIPW32+1.96*SE.ATEIPW32.bs
mean(ifelse(LCLIPW32<tau1&UCLIPW32>tau1, 1,0))

LCLSATE21<-SATE21-1.96*SE.ATESTRAT21.bs
UCLSATE21<-SATE21+1.96*SE.ATESTRAT21.bs
mean(ifelse(LCLSATE21<tau1&UCLSATE21>tau1, 1,0))

LCLSATE31<-SATE31-1.96*SE.ATESTRAT31.bs
UCLSATE31<-SATE31+1.96*SE.ATESTRAT31.bs
mean(ifelse(LCLSATE31<tau2&UCLSATE31>tau2, 1,0))

LCLSATE32<-SATE32-1.96*SE.ATESTRAT32.bs
UCLSATE32<-SATE32+1.96*SE.ATESTRAT32.bs
mean(ifelse(LCLSATE32<tau1&UCLSATE32>tau1, 1,0))

LCLATEDR21<-ATEDR21-1.96*SE.ATEDR21.bs
UCLATEDR21<-ATEDR21+1.96*SE.ATEDR21.bs
mean(ifelse(LCLATEDR21<tau1&UCLATEDR21>tau1, 1,0))

LCLATEDR31<-ATEDR31-1.96*SE.ATEDR31.bs
```

```
UCLATEDR31<-ATEDR31+1.96*SE.ATEDR31.bs
mean(ifelse(LCLATEDR31<tau2&UCLATEDR31>tau2, 1,0))

LCLATEDR32<-ATEDR32-1.96*SE.ATEDR32.bs
UCLATEDR32<-ATEDR32+1.96*SE.ATEDR32.bs
mean(ifelse(LCLATEDR32<tau1&UCLATEDR32>tau1, 1,0))

LCLATEDR21co<-ATEDR21co-1.96*SE.ATEDR21co.bs
UCLATEDR21co<-ATEDR21co+1.96*SE.ATEDR21co.bs
mean(ifelse(LCLATEDR21co<tau1&UCLATEDR21co>tau1, 1,0))


LCLATEDR31co<-ATEDR31co-1.96*SE.ATEDR31co.bs
UCLATEDR31co<-ATEDR31co+1.96*SE.ATEDR31co.bs
mean(ifelse(LCLATEDR31co<tau2&UCLATEDR31co>tau2, 1,0))


LCLATEDR32co<-ATEDR32co-1.96*SE.ATEDR32co.bs
UCLATEDR32co<-ATEDR32co+1.96*SE.ATEDR32co.bs
mean(ifelse(LCLATEDR32co<tau1&UCLATEDR32co>tau1, 1,0))
```

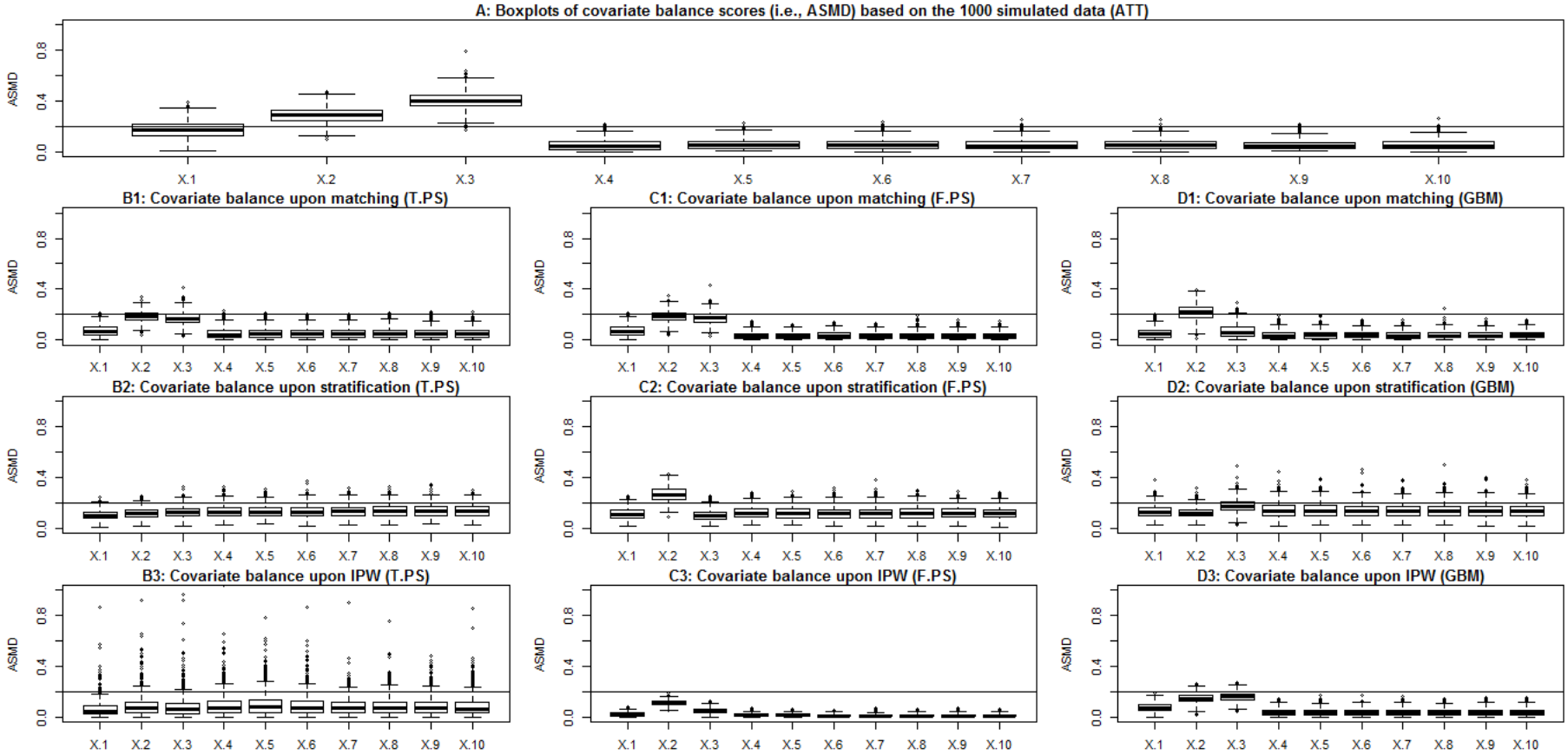

Figure A1: Assessment of covariates balancing of different methods for estimating ATT based on 1000 simulated data with a sample size of 1000, where the covariates are independently normally distributed. Panel A is the boxplot of the balancing scores (i.e., ASMD) for each covariate; Panels B1-B3 are the boxplots of the balancing score for each covariate upon matching, stratification, and IPW, respectively, when propensity score is estimated using the true logistic regression model; Panels C1-C3 are the boxplots of the balancing scores of each covariate when propensity score is estimated using the false logistic regression model; Panels D1-D3 are the boxplots of the balancing scores for each covariate when propensity score is estimated using the GBM model.

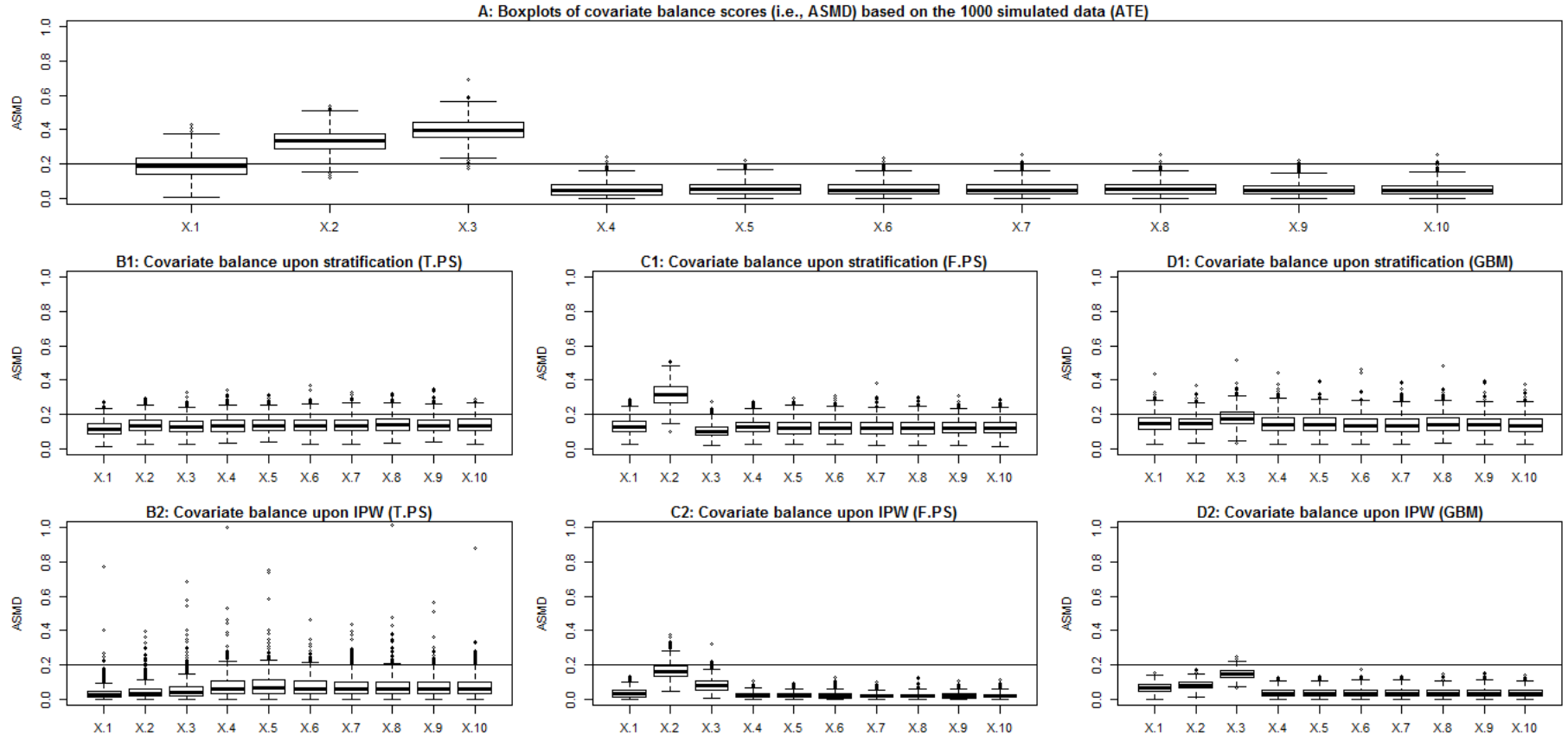

Figure A2: Assessment of covariates balancing of different methods for estimating ATE based on 1000 simulated data with a sample size of 1000, where the covariates are independently normally distributed. Panel A is the boxplot of the balancing scores (i.e., ASMD) for each covariate; Panels B1-B2 are the boxplots of the balancing scores for each covariate when propensity score is estimated using the true logistic regression model; Panels C1-C2 are the boxplots of the balancing scores for each covariate when propensity score is estimated using the false logistic regression model; Panels D1-D2 are the boxplots of the balancing scores for each covariate when propensity score is estimated using the GBM model.

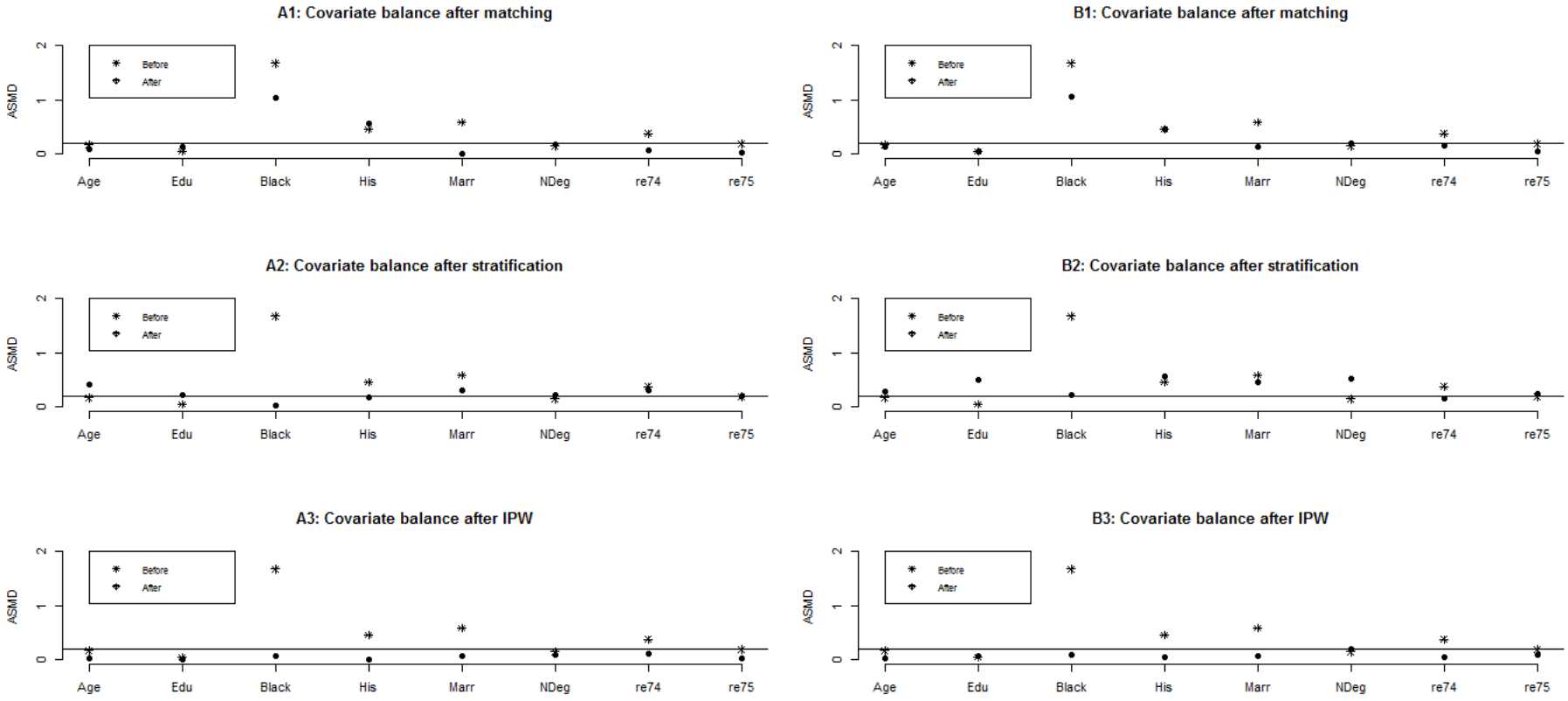

Figure A3: Assessment of covariates balancing for the Lalonde study, when estimating ATT. Panel A1 - A3 are the graphs for the covariate balance after matching, stratification, and IPW respectively when propensity score is estimated using the logistic regression model. Panel B1 - B3 are the graphs for the covariate balance after matching, stratification, and IPW respectively when propensity score is estimated using the GBM model.

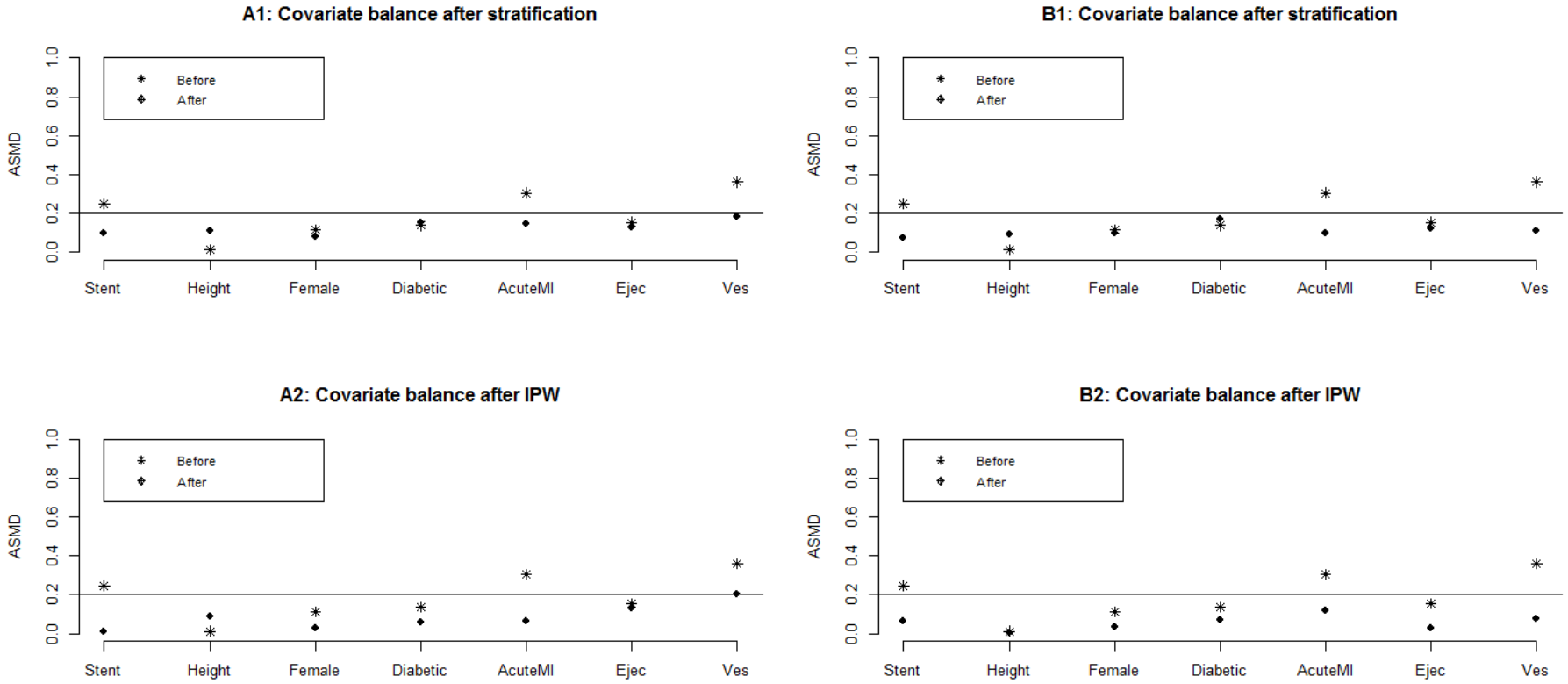

Figure A4: Assessment of covariates balancing for the Lindner study, when estimating ATE. Panel A1 - A2 are the plots for the covariate balance before and after stratification and IPW, when propensity score is estimated using the logistic regression model. Panel B1 - B2 are the plots for the covariate balance before and after stratification and IPW, when propensity score is estimated using the GBM model.

CURRICULUM VITAE

NAME: Younathan Abdia

ADDRESS: Department of Biostatistics and Bioinformatics
University of Louisville
Louisville, KY 40292

EDUCATION: Master of Sceince, The Ohio State University, OH, 2011
Master of Public Health, Saint Louis University, MO, 2009

PROFESSIONAL EXPERIENCE: Graduate Research Assistant, University of Louisville
Aug 2012 - Present

Lecturer, Forman Christian College, Lahore, Oct 2011 to July 2012

Adjunct Faculty, LUMS, Lahore, Jan 2012 to July 2012

PUBLICATIONS: **Abdia, Y.***, Kulasekera, K. B., Datta, S., and Kong, M. Estimating average treatment effects among multiple treatment groups by using adaptive ensemble method. (In Progress)

Abdia, Y.*, Kulasekera, K. B., Datta, S., Boakye, M., and Kong, M. Propensity score based methods for estimating average treatment effect and average treatment effect among treated: A comparative study. Preprint

Sikdar, S., Wosoba, H.C., **Abdia, Y**., Dutta, S., Gill, R., Datta, S., Datta, S. An Intergrative exploratory analysis of -omics data for the ICGC cancer genomes lung adenocarcinoma study. *System Biomedicine* 2, 56-64 **2014**

PRESENTATIONS

ORAL

Propensity score based methods for estimating treatment effect, Spring 2016 Seminar Series, Department of Bioinformatics and Biostatistics, University of Louisville, January 15, 2016.